Polygenic architecture informs potential vulnerability to drug-induced liver injury

Authors and affiliations:

Masaru Koido[1,2,#], Eri Kawakami[2,3,#], Junko Fukumura[1,2], Yui Noguchi[1,2], Momoko Ohori[2,3], Yasunori Nio[2,3], Paola Nicoletti[4], Guruprasad P. Aithal[5], Ann K. Daly[6], Paul B. Watkins[7], Hisashi Anayama[8], Yvonne Dragan[9], Tadahiro Shinozawa[2,3,8] and Takanori Takebe[1,2,3,10,11,12,13]*

1. Department of Regenerative Medicine, Yokohama City University Graduate School of Medicine, Kanazawa-ku 3-9, Yokohama, Kanagawa 236-0004, Japan.

2. Organoid medicine project, T-CiRA joint program, Fujisawa, Kanagawa 251-8555, Japan

3. Regenerative Medicine Unit, Takeda Pharmaceutical Company Ltd, Fujisawa, Kanagawa 251-8555, Japan

4. Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, USA

5. NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and the University of Nottingham, Nottingham, UK.

6. Translational & Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK.

7. Institute for Drug Safety Sciences, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, North Carolina, USA

8. Drug Safety Research and Evaluation, Takeda Pharmaceutical Company Ltd, Fujisawa, Kanagawa 251-8555, Japan

9. Global Discovery and Investigative Toxicology, Takeda Pharmaceuticals, 35 Landsdowne St, Cambridge, MA 02139, USA

10. Institute of Research, Tokyo Medical and Dental University 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8510, Japan

11. Division of Gastroenterology, Hepatology and Nutrition & Division of Developmental Biology, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, OH 45229-3039, USA

12. The Center for Stem Cell and Organoid Medicine (CuSTOM), Cincinnati Children’s Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, OH 45229- 3039, USA.

13. Department of Pediatrics, University of Cincinnati College of Medicine, 3333 Burnet Avenue, Cincinnati, OH 45229-3039, USA

[#] These authors contributed equally to this work.

*Corresponding author: Takanori Takebe, E-mail: Takanori.Takebe@cchmc.org, Phone: +1-513-803-7807

Abstract

Drug-Induced-Liver-Injury (DILI) is a leading cause of termination in drug development programs and removal of drugs from the market, and this is partially due to the inability to identify patients who are at risk[1]. Here, we developed a polygenic risk score (PRS) for DILI by aggregating effects of numerous genome-wide loci identified from previous large-scale genome-wide association studies (GWAS)[2]. The PRS predicted the susceptibility to DILI in patients treated with fasiglifam, amoxicillin-clavulanate or flucloxacillin, and in primary hepatocytes and stem cell-derived organoids from multiple donors treated with over 10 different drugs. Pathway analysis highlighted processes previously implicated in DILI, including unfolded protein responses and oxidative stress. In silico screening identified compounds that elicit transcriptomic signatures present in hepatocytes from individuals with elevated PRS, supporting mechanistic links and suggesting a novel screen for safety of new drug candidates. This genetic-, cellular-, organoid- and human-scale evidence underscored the polygenic architecture underlying DILI vulnerability at the level of hepatocytes, thus facilitating future mechanistic studies. Moreover, the proposed “polygenicity-in-a-dish” strategy might potentially inform designs of safer, more efficient, and robust clinical trials.

Main

Polygenic risk scores encompassing cumulative effects of tens of thousands to hundreds of thousand variants or genome-wide variants, can identify individuals at high risk of common diseases[3]. However, it remains unclear whether a polygenic model can account for susceptibility to relatively rare diseases such as drug-induced liver injury (DILI).

DILI is one of the leading causes of termination in drug development programs and drug withdrawals from the market; in clinical practice, it is also associated with substantial morbidity and mortality[1]. Genome-wide association studies (GWAS) conducted by the international Drug-Induced Liver Injury Consortium (iDILIC) and the Drug-Induced Liver Injury Network (DILIN) have identified several significant variants associated with DILI due to multiple different drugs (*P*-value from GWAS (P_{GWAS}) < 5×10^{-8})[2]. Since each variant has modest predictive impact, we herein revisited GWAS findings to determine whether the polygenic score, which sums up effects of numerous variants, informs potential DILI susceptibility in humans. The goal of this study was to develop a polygenic risk scores (PRS) using data from previous GWAS (performed by iDILIC and DILIN) and to validate their performance characteristics in GWAS data obtained from an independent clinical trial of a hepatotoxic drug, as well as multiple

donor-derived organoids and primary hepatocytes treated with a variety of hepatotoxic medications. We also used the derived PRS to identify mechanisms underlying the susceptibility to DILI.

To assess the polygenicity associated with susceptibility to DILI, summary statistics of a GWAS from the iDILIC/DILIN collaboration which studied cases of DILI due to selected drugs were used for the discovery dataset, including all patients (ALL-DILI GWAS, n=862), hepatocellular (HC-DILI GWAS, n=474) and cholestatic/mixed (CM-DILI GWAS, n=323) and compared with those of population-matched controls (n=10,588) (Figure 1a and Extended Data Fig.1; see details in ref#[2]). In the original report by iDILIC/DILIN, two loci in the entire cohort (ALL-DILI), one locus in CM-DILI GWAS and a locus in HC-DILI GWAS, were genome-wide significant ($P_{GWAS} < 5\times10^{-8}$)[2]. Herein, we hypothesized that polygenic architecture informed by iDILIC/DILIN GWAS would predict the susceptibility of an independent DILI GWAS dataset, namely patients who experienced liver injury during the clinical trials of TAK-875 (fasiglifam). As linkage disequilibrium (LD) score regression is inappropriate with the relatively small sample size in GWAS study[4], we sought to investigate the polygenic architecture of DILI susceptibility by comparing two independent DILI GWAS.

TAK-875 is an oral, highly potent, and selective FFAR1 (GPR40) agonist, the first of its class tested for glucose-lowering ability in type 2 diabetes patients[5]. TAK-875 development was discontinued owing to the rare DILI incidence detected in a Phase 3 study[6]. We recruited 43 TAK-875 DILI patients (cases) and 129 matched patients who did not experience liver injury when treated with TAK-875 in the clinical trials (controls) for subsequent GWAS. The TAK-875 cases and controls were not included in the iDILIC/DILIN GWAS. Although an increase in ALT levels was observed after treatment with TAK-875 in the DILI patients, total bilirubin was also increased (Extended Data Fig.2) consistent with experimental models that have demonstrated an inhibition of bile acid transporters by TAK-875 treatment[7]. After QC analysis and genome-wide genotype imputation using populations from 1000 Genomes Phase 3 (1KGP3) as a reference, 39 TAK-875 DILI cases and 122 controls who were self-reported Caucasians with 13,477,278 autosomal variants and 184,010 non-imputed variants were selected for further study (Extended Data Fig.1 and 2c-e; Supplementary Table S1). Logistic regression analysis with covariates for population structure yielded no genome-wide significant association signals in the GWAS (Figure 1b).

After iDILIC/DILIN GWAS summary statistics were clumped by linkage disequilibrium (LD) in the European population, we performed shared genetic aetiology

112 analysis to determine whether the iDILIC/DILIN GWAS signals correlated with TAK-875
113 DILI GWAS signals using maximal inclusion of single nucleotide polymorphisms (SNPs)
114 with $P_{GWAS}<0.5$ (Figure 1c). Interestingly, comparing GWAS results showed that
115 iDILIC/DILIN GWAS for CM-DILI (Figure 1c) shared genetic aetiology with TAK-875 DILI
116 more than did the iDILIC/DILIN GWAS for ALL- or HC-DILI did (Extended Data Fig.2f,g),
117 indicating that susceptibility to TAK-875 DILI is genetically linked to CM-DILI
118 susceptibility to many other drugs.

119 Functional enrichment analysis of summary statistics for iDILIC/DILIN CM-DILI
120 was performed using GARFIELD (GWAS analysis of regulatory or functional information
121 enrichment with LD correction) software using default annotations including >1000
122 publicly available datasets with regulatory genomic regions [8]. The strongest enrichment
123 in the epigenetic state was observed in HepG2 cells vs cells of other origin (Figure 1d,
124 left), suggesting that a substantial genetic susceptibility for CM-DILI resides within the
125 hepatocyte-like cellular context. The strongest enrichment in a genic state was in the 5'
126 untranslated region (UTR) (Figure 1d, right). The Pascal (pathway scoring algorithm)[9]
127 revealed that association signals of CM-DILI GWAS were enriched in experimentally
128 proposed pathways for DILI, including mitochondrial activity, oxidant-induced survival[10],
129 and unfolded protein response (UPR)[11] ($p<0.05$, Figure 1e and Supplementary Table

S2). Thus, genetic annotation of genome-wide CM-DILI GWAS signals reasonably reflect reported molecular and cellular pathways related to human DILI. We hereafter defined the CM-PRS_{gw} (genome-wide PRS) by aggregating 27,740 SNPs in and near hepatocyte-expressed genes with $P_{GWAS}<0.5$ for further analysis (see Methods).

Although CM-PRS_{gw} is solely formed by a iDILIC/DILIN study that does not include TAK-875 subjects, the TAK-875 DILI case samples had a marginally higher CM-PRS_{gw} score distribution than control samples (AUROC 0.61 (95% CI: 0.51-0.71), *P* for two-tailed Wilcoxon–Mann–Whitney U test<0.05) (Figure 1f and Extended Data Fig.3). Interestingly, there was a trend towards an increase in the baseline values of ALT, AST, and total bilirubin as the CM-DILI PRS_{gw} increased in all TAK-875 treated patients (Extended Data Fig.4). We also found that CM-DILI PRS_{gw} potentially predicted susceptibility to DILI of 167 flucloxacillin or 207 amoxicillin-clavulanate patients with a CM-DILI phenotype (AUROC 0.61 (95%CI: 0.57–0.66) or 0.62 (0.58–0.66), respectively; Extended Data Fig.5; see Supplementary Note and Supplementary Methods). These cases were not included in the original CM-DILI PRS determinations[2] but described elsewhere (see Cirulli et al[12]). These data support that the cholestatic/mixed polygenic score (CM-PRS_{gw}) could predict the potential DILI susceptibility of patients caused by multiple different drugs. Notably, when CM-DILI

148 GWAS added 157 flucloxacillin and 320 amoxicillin-clavulanate DILI patients (Figure 1a),
149 the GWAS signals were also enriched in plausible biological pathways and
150 hepatocyte-like context (Extended Data Fig.6, and Supplementary Table S3), and its
151 derived PRS, named CM-PRS$_{gw+}$, was comparable to the above in terms of
152 distinguishing TAK-875 DILI cases from controls (Figure 1f and Extended Data Fig.3;
153 see Supplementary Note).

154 Next, we investigated whether the two different PRSs could stratify DILI vulnerability
155 under CM-DILI conditions using both a scored human hepatocyte model and human
156 liver bud organoids (HLOs) derived from iPSC (Figure 2a). Primary human hepatocytes
157 (PHH), the gold standard for toxicology testing (and a fully matured adult cell type) were
158 randomly selected from vendors (Supplementary Table S4). Multi-donor iPSCs were
159 obtained from the cell bank, Coriell Institute and the European Bank for Induced
160 Pluripotent Stem Cells (EBiSC) (Supplementary Table S5). Whole-genome genotypes
161 of purchased PHH and iPSCs from healthy European (EUR) donors were determined
162 by SNP array to calculate their CM-PRS$_{gw}$ and CM-PRS$_{gw+}$ (see Methods). PHH (21
163 donors) and iPSC-HLOs (5 donors) in this study were selected after a battery of quality
164 control experiments (Extended Data Fig.7 and 8). For the modelling the CM- DILI
165 phenotypes with the iPSCs, we adopted modified protocols of previously reported

166 vascularized 3D iPSC-HLOs[13]. All the donor iPSC successfully generated HLOs with
167 consistent expression of liver-associated proteins (Albumin and bile salt export pump;
168 BSEP), and CYP3A4 activity (Extended Data Fig.7a-d). HLOs that were exposed to
169 cholyl-lysyl-fluorescein (CLF), a fluorescently labelled bile acid, displayed prominent
170 accumulation of bile acids under cyclosporin A (CsA) treatment, followed by massive
171 cell death (Extended Data Fig.7e, f). Hereafter we used this as a model for assessing
172 CM-DILI by iPSC-HLOs. For the modelling of the CM-DILI phenotypes from PHH cells,
173 we analysed the viability based on established cholestatic-type DILI protocols[14] by
174 pre-treatment with lithocholic acid (LCA), a highly toxic secondary bile acid (Extended
175 Data Fig. 9 a-f).

176 Under the cholestatic conditions, we evaluated the toxicity of 12 DILI drugs,
177 cyclosporine, bosentan, troglitazone, diclofenac, flutamide, ketoconazole,
178 carbamazepine, amoxicillin-clavulanate and methapyrilene (drugs that produce DILI
179 with a CM phenotype), tacrine, acetaminophen (APAP), tolcapone (drugs that produced
180 DILI with an hepatocellular phenotype), to determine if the score-dependent trend is
181 conserved across multiple drugs (Figure 2a). The correlation patterns were plotted
182 between the two different PRS and cell viability for each drug treatment (Figure 2b).
183 CM-PRS_{gw} and the cell death rate under drug treatment were well correlated, and the

184 average Pearson correlation coefficient of each category of drugs used in the
185 experiment was more significant than that of CM-PRS$_{gw+}$(Figure 2c), Thus, by
186 examining the correlation comparison using iPSC-HLOs and PHH cells, we found that
187 CM-PRS$_{gw}$ is more strongly correlated with the toxicity evaluation of the DILI drugs than
188 CM-PRS$_{gw+}$ (Figure 2b and 2c). Intriguingly, the five-donor iPSC-HLO without LCA
189 pre-treatment also revealed a correlation between cell viability under CsA treatment and
190 the CM-PRS$_{gw}$ ($P < 0.05$), whereas CM-PRS$_{gw+}$ showed no obvious correlation
191 (Extended Data Fig. 9g,h). CM-PRS$_{gw+}$ was developed from GWAS of our original
192 cohort plus 477 patients with DILI due to either amoxicillin-clavulanate or flucloxacillin
193 so that almost half of the cases used to derive CM-PRS$_{gw+}$ were just due to these two
194 drugs. The fall-off in prediction suggests that CM-PRS$_{gw+}$ may not be able to identify
195 DILI liability for all drugs in our experimental models. Taken together, these data show
196 that the CM-DILI PRS$_{gw}$-based stratification approach in the multi-donor PHH- and
197 iPSC-HLO-based assay was more widely applicable than PRS$_{gw+}$-based one.

198 To approach functional pathways associated with CM-DILI risks, we first performed
199 comparative transcriptome analysis of basal expression levels in the cells in culture,
200 regressing out five transcriptome principal components to capture experimental
201 variability[15] (Figure 3a). Unbiased pathway enrichment analysis in 10 PHH donors

revealed that genes involved in mitochondria and translation were inactivated in the higher CM-PRS_{gw} donors (FDR<0.01; Figure 3b and Supplementary Table S6). We next determined the CM-PRS_{gw} in 157 donors of human liver samples that have undergone RNA profiling[16]. There too, the donors with the highest CM-PRS_{gw} scores appeared to have the same pathways inactivated as we observed in the 10 PHH donors (Figures 3a, 3b, and Supplementary Table S7; see Methods). In addition, basal expression levels of mitochondrial genes and UPR-related genes were correlated with the CM-DILI PRS_{gw} in the transcriptome analysis of 5 donor iPSC-HLOs (Figures 3c). These data suggest that global transcriptional and translational processes might be perturbed in the people who are susceptible to CM-DILI, consistent with findings of GWAS-signal enrichments in 5'UTR region (Figure1e).

Based on these pathway extractions, we further investigated CM-DILI vulnerability mechanisms using an iPSC-HLO model focusing on mitochondrial activity and ER stress signals under CsA treatment. Consistently, we observed cholestatic cell death with elevated UPR-associated proteins, massive ROS production and functional mitochondria depletion in iPSC-HLO (Figure 3d-f and Extended Data Fig. 10). In agreement with this, an antioxidant GSH/GSSG ratio was elevated more under CM-DILI drug treatment in a CM-PRS_{gw} higher donor than the lower donor (Figure 3g), and this

220 was partially alleviated by Bardoxolone Methyl (BM), a potent activator of Nuclear factor
221 like-2 factor[17] (Figure 3h). The enhanced antioxidant response induced by BM protects
222 against hepatocyte death in cholestatic conditions (Figure 3i) even in the CM-PRS_{gw}
223 higher donor. Collectively, ROS activation may be a potential downstream event prior to
224 hepatocyte death owing to the accumulation of bile acids and a predisposition to
225 transcriptional and translational stress substantiated by PRS.

226 To verify the connection between CM-DILI vulnerable pathways and CM-DILI
227 drugs, we analysed the risk score associated transcriptomic signatures in PHH from 8
228 donors under CsA treatment (see Supplementary Table S4) and compared the
229 transcriptome data in Toxicogenomics Project-Genomics Assisted Toxicity Evaluation
230 system (TG-GATEs)[18] including 941 distinct conditions of drugs-treatment related
231 changes in transcriptome in PHH (Figure 4a and Supplementary Table S8). First, the
232 genes highly inactivated in higher CM-PRS_{gw} donors' PHH were enriched (FDR<0.01) in
233 pathways such as respiratory electron transport (*i.e.,* inducible mitochondrial toxicity),
234 translational regulation, and mRNA processing (Figure 4b and Supplementary Table S9),
235 in concordance with baseline transcriptome analysis in Figure 3. To extend these
236 findings, we searched for the drugs which inactivated these DILI vulnerable pathways in
237 TG-GATEs (FDR < 0.01; Supplementary Table S10). Many well-known CM-DILI drugs,

238 such as CsA, flutamide, ketoconazole, ranitidine, sulpiride, valproic acid, azathioprine,
239 decreased levels of genes in these DILI vulnerable pathways (Figure 4c), supporting our
240 mechanistic findings with CM-DILI inducible drugs in *in vitro* multi-donor PHH analysis
241 (Supplementary Table S10). We note that UPR-related genes were also activated in the
242 higher CM-DILI PRS_{gw} PHH (Figure 4d) as shown in Figure 3. These data suggest that
243 *in silico* toxicity modelling, coupled with *in vitro* genomic and transcriptomic approaches,
244 may identify compounds with potential for DILI before entering clinical trials.

245 This is the first unbiased analysis to implicate genetic variation at the level of
246 the hepatocyte contributing to DILI susceptibility, validated in an independent clinical
247 study as well as unrelated donor-derived organoids and primary hepatocytes.
248 Hepatocyte damage due to an imbalance between formation and clearance of reactive
249 metabolites, steps that are common for many compounds, is likely upstream events
250 mediating liver injury[19]. Once hepatocytes are damaged, innate and adaptive immunity
251 play a significant role in driving tissue inflammation and injury[20], as supported by
252 identification of HLA alleles and haplotypes as DILI risk factors in GWAS
253 (Supplementary Note). However, while some of the cases in the iDILIC/DILIN cohort
254 were shown to have a significant HLA association[2], the overall HLA contribution in the
255 entire cohort is relatively limited. This is similar to TAK-875 DILI, where no HLA

association has been detected. The PRS (PRS_{gw+}) derived from the expanded iDILIC/DILIN cohort included mostly flucloxacillin and amoxicillin-clavulanate cases, for which stronger HLA associations have been demonstrated [21,22] was still useful in predicting CM-DILI, but not as well as PRS_{gw}. We speculate that this poorer performance may be due to limited contribution of non-HLA risk factors when HLA is the main genetic risk factor. Up to the present, few non-HLA genetic risk factors for any form of DILI have been identified and replicated, thus highlighting the importance of aggregating many variants, each making a small contribution to overall DILI risk.

Our PRS_{gw} for CM-DILI revealed shared DILI predictivity across a variety of drugs independent of their individual chemical characteristics that are considered important[23]. This indicates that the make-up of our polygenic scores relates to intra-hepatocyte mechanisms leading to hepatotoxicity. For example, both UPR and oxidative stress have been reportedly associated with DILI pathogenesis and cause cell death due to experimental cholestasis using various drugs such as flucloxacillin, levofloxacin, diclofenac and carbamazepine[24,25]. Consistently, our *in vitro* polygenic score-based approach, coupled with genomic, transcriptomic, and phenotypic approaches, indicated that diverse biological pathways including UPR and oxidative stress responses in hepatocytes contribute to CM-DILI susceptibility (Figure 4e). These

findings can also provide fundamental mechanisms defining CM-DILI and may ultimately contribute to the design of safer, efficient, and more robust clinical trials. More broadly, the proposed “polygenicity-in-a-dish strategy” is a powerful approach to investigate and interrogate highly complicated pathogenesis in humans with minimal confounding factors.

References

1. Chalasani, N. *et al.* Features and Outcomes of 899 Patients With Drug-Induced Liver Injury: The DILIN Prospective Study. *Gastroenterology* **148**, 1340-1352.e7 (2015).

2. Nicoletti, P. *et al.* Association of Liver Injury From Specific Drugs, or Groups of Drugs, With Polymorphisms in HLA and Other Genes in a Genome-Wide Association Study. *Gastroenterology* **152**, 1078–1089 (2017).

3. Khera, A. V. *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).

4. Bulik-Sullivan, B. *et al.* LD score regression distinguishes confounding from

polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

5. Burant, C. F. *et al.* TAK-875 versus placebo or glimepiride in type 2 diabetes mellitus: a phase 2, randomised, double-blind, placebo-controlled trial. *Lancet* **379**, 1403–1411 (2012).

6. Marcinak, J. F., Munsaka, M. S., Watkins, P. B., Ohira, T. & Smith, N. Liver Safety of Fasiglifam (TAK-875) in Patients with Type 2 Diabetes: Review of the Global Clinical Trial Experience. *Drug Saf.* **41**, 625–640 (2018).

7. Wolenski, F. S. *et al.* Fasiglifam (TAK-875) alters bile acid homeostasis in rats and dogs: A potential cause of drug induced liver injury. *Toxicol. Sci.* **157**, 50–61 (2017).

8. Iotchkova, V. *et al.* GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nat. Genet.* **51**, 343–353 (2019).

9. Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z. & Bergmann, S. Fast and Rigorous Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. *PLoS Comput. Biol.* **12**, 1–20 (2016).

10. Kass, G. E. N. & Price, S. C. Role of Mitochondria in Drug-Induced Cholestatic Injury. *Clin. Liver Dis.* **12**, 27–51 (2008).

11. Vatakuti, S., Olinga, P., Pennings, J. L. A. & Groothuis, G. M. M. Validation of precision-cut liver slices to study drug-induced cholestasis: a transcriptomics approach. *Arch. Toxicol.* **91**, 1401–1412 (2017).

12. Cirulli, E. T. *et al.* A Missense Variant in PTPN22 is a Risk Factor for Drug-induced Liver Injury. *Gastroenterology* **156**, 1707-1716.e2 (2019).

13. Takebe, T. *et al.* Massive and Reproducible Production of Liver Buds Entirely from Human Pluripotent Stem Cells. *Cell Rep.* **21**, (2017).

14. Ogimura, E., Sekine, S. & Horie, T. Bile salt export pump inhibitors are associated with bile acid-dependent drug-induced toxicity in sandwich-cultured hepatocytes. *Biochem. Biophys. Res. Commun.* **416**, 313–317 (2011).

15. Delaneau, O. *et al.* Chromatin three-dimensional interactions mediate genetic effects on gene expression. *Science* **364**, eaat8266 (2019).

16. Schadt, E. E. *et al.* Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* **6**, 1020–1032 (2008).

17. Hybertson, B. M., Gao, B., Bose, S. K. & McCord, J. M. Oxidative stress in health

and disease: The therapeutic potential of Nrf2 activation. *Mol. Aspects Med.* **32**, 234–246 (2011).

18. Igarashi, Y. *et al.* Open TG-GATEs: a large-scale toxicogenomics database. *Nucleic Acids Res.* **43**, D921-7 (2015).

19. Kaliyaperumal, K. *et al.* Pharmacogenomics of drug-induced liver injury (DILI): Molecular biology to clinical applications. *Journal of Hepatology* vol. 69 948–957 (2018).

20. Chen, M., Suzuki, A., Borlak, J., Andrade, R. J. & Lucena, M. I. Chen M, Suzuki A, Borlak J, Andrade RJ, Lucena MI. Drug-induced liver injury: Interactions between drug properties and host factors. Journal of hepatology. 2015 Aug 31;63(2):503-14. *Journal of Hepatology* vol. 63 503–514 (2015).

21. Daly, A. K. *et al.* HLA-B*5701 genotype is a major determinant of drug-induced liver injury due to flucloxacillin. *Nat. Genet.* **41**, 816–819 (2009).

22. Lucena, M. I. *et al.* Susceptibility to amoxicillin-clavulanate-induced liver injury is influenced by multiple HLA class i and II alleles. *Gastroenterology* **141**, 338–347 (2011).

23. European Association for the Study of the Liver. Electronic address:

easloffice@easloffice.eu, Clinical Practice Guideline Panel: Chair:, Panel members & EASL Governing Board representative: EASL Clinical Practice Guidelines: Drug-induced liver injury. *J. Hepatol.* **70**, 1222–1261 (2019).

24. Fredriksson, L. *et al.* Drug-induced endoplasmic reticulum and oxidative stress responses independently sensitize toward TNFα-mediated hepatotoxicity. *Toxicol. Sci.* **140**, 144–59 (2014).

25. Burban, A., Sharanek, A., Guguen-Guillouzo, C. & Guillouzo, A. Endoplasmic reticulum stress precedes oxidative stress in antibiotic-induced cholestasis and cytotoxicity in human hepatocytes. *Free Radic. Biol. Med.* **115**, 166–178 (2018).

Methods

iDILIC/DILIN GWAS summary statistics

iDILIC/DILIN GWAS summary statistics (SNP name, minor allele, odds ratio for the minor allele, and p-value) were kindly provided by researchers in these consortia[2]. GWAS were conducted using three types of cohorts, DILI patients (ALL-DILI, n=862), those with hepatocellular injury (HC-DILI, n=474), and those with cholestatic/mixed injury (CM-DILI, n=323), against population-matched controls (n=10,588). We used biallelic and unambiguous SNPs catalogued in snpdb147 and included in the 1000 genome project phase 3 (1KGP3) imputation reference panel (Minimac3 website) [26]. The odds ratios of iDILIC/DILIN GWAS summary statistics were aligned towards the alternative allele in the reference panel. Finally, 4,392,401 SNPs in ALL-DILI, 4,392,226 SNPs in CM-DILI, and 4,393,472 SNPs in HC-DILI were selected (referred to as processed iDILIC/DILIN GWAS summary statistics).

TAK-875 patients and controls collection criteria

Patients comprised individuals treated with TAK-875 monotherapy or combinatorial therapy. Case (DILI) were identified as subjects receiving TAK-875 and experiencing within 7 days post treatment a rise in serum Alanine Aminotransferase (ALT) or

Aspartate Aminotransferase (AST) at least ≥3× upper limits of normal and at least 2x their baseline value. Otherwise eligible patients who first met the DILI inclusion criteria more that 7 after final dosing were excluded from the analysis. Control subjects were randomly identified and did not experience ALT/AST elevations exceeding 10% from baseline (and not exceeding 3× ULN). Controls were matched in a 3:1 ratio to cases on the basis of study, drug treatment, race, and sex.

TAK-875 GWAS

The study population of the TAK-875 GWAS included subjects experiencing DILI (cases) and corresponding matched controls as defined above. Data regarding the patients and controls were collected from TAK-875 Phase 2 and 3 clinical studies. All studies were conducted in accordance with the Declaration of Helsinki, International Conference on Harmonization Guidelines for Good Clinical Practice, and all applicable local regulatory requirements. All studies including protocol and informed consent forms were approved by the Institutional Review Board (IRB) at each study site. The IRB or independent ethics committee for each site was constituted according to the applicable requirements of the participating region, approving the protocol and subject informed consent. Prior to undergoing any study procedures, signed and dated informed consent

389 form was received from each patient.

390 We included self-reported Caucasian individuals (Case *n*=43, Ctrl *n*=129) for
391 the GWAS. These subjects were genotyped in a combination of HumanOmni5-Quad
392 BeadChip (Illumina) and Infinium Human Exome-12 v1.2 BeadChip (Illumina).
393 Corresponding manifest files of genotyping arrays, HumanOmni5-4v1_B, and
394 HumanExome-12-v1-2-B, respectively, were used. Quality control (QC) analysis of the
395 genotyped data checked sample and variant call rate, probe-target duplication,
396 missingness, heterozygosity, Hardy–Weinberg equilibrium, identity by descent,
397 ancestors, and minor allele frequency (MAF) using plink software (version v1.90b3.44)[27]
398 (see Supplementary Methods). The haplotype phase was estimated using Eagle
399 software (version 2.3)[28] with the 1KGP3 reference panel (provided in Minimac3 website,
400 version 5, *n*=2,504). Genotypes of autosomal variants were imputed using Minimac3
401 software (version 2.0.1) with the 1KGP3 imputation reference panel[29]. Finally, Case
402 *n*=39 and Ctrl *n*=122 samples with autosomal 13,477,278 variants with a minimac3
403 imputation quality (Rsq) ≥ 0.7 and non-imputed genotyped 184,010 variants were
404 included in the GWAS. GWAS was performed using imputed allele dosages of variants
405 with a MAF > 1% and fitted to an additive genetic model of logistic regression with PC1–
406 PC5 in white GWAS, using snptest software (version 2.5.2)[30]. A Manhattan plot was

407 performed using R package *qqman* v0.1.2.

408

409 **Genetic analysis of shared aetiology**

410 We evaluated shared genetic aetiology between two GWAS summary statistics
411 using PRSice software (version 1.25)[31]. Common autosomal SNPs between the
412 TAK-875 GWAS panel and the processed iDILIC/DILIN GWAS summary statistics were
413 used. Base summary statistics were pruned from LD-based clumps with parameters
414 recommended in the PRSice software: clump.p1 0.5, clump.p2 0.5, and clump.kb 300
415 using plink software (version v1.90b3.44)[27]. In this clumping step, 1KGP3 European
416 super population was used and 51,052 SNPs for ALL-DILI, 50,029 SNPs for CM-DILI,
417 and 50,553 SNPs for HC-DILI were selected. SNPs with OR≠1 in target GWAS sets
418 were selected. If the β-value for the association of SNPs in TAK-875 GWAS was
419 positive infinity (negative infinity), secondary maximum (or minimum) β was replaced
420 with it. Standard errors of datasets were calculated from OR and *p*-values, using Wald
421 statistics. Finally, variance explained by the base risk score in target GWAS sets and
422 *p*-values were obtained from the PRSice software[31]. We noted that the increasing trends
423 are important to show the polygenic architecture in DILI and the levels of variance
424 explained might be high due to the winner's curse issue. To avoid incorrect conclusions

425 from genetic overlap due to population stratification problem, we only conducted the
426 summary statistics-based analysis.

427

428 **PRS analysis**

429 From the processed iDILIC/DILIN GWAS summary statistics for the European
430 LD-based clumped data, we further selected imputable SNP sets from the 1KGP3
431 reference panel. PRS was calculated as follows: (β for association) $\times$ (number of effect
432 allele for β). The number of alleles was based on the best-guessed genotype upon
433 1KGP3 imputation with Rsq≥0.7.

434 To determine a PRS_{gw}, we considered the cellular context of hepatocytes to
435 observe the phenotype. Hence, we obtained RNA sequencing (RNA-seq) data of public
436 primary human hepatocytes (SRR4000958)[32]. Reads from the FastQ files were mapped
437 to the GRCh38.p10 reference sequence, using GENCODE v26 annotation file and
438 STAR software (version 2.5.2b)[33], after quality control by FaQCs software (version
439 1.34)[34] with default parameters. FPKM values were obtained using cufflinks software
440 (version 2.2.1)[35], wherein the max-bundle-frags option was adjusted to 10^9
441 fragments/locus. Expressed genes (n=14,576) were considered to have an FPKM>0 in
442 the PHH. Regions 10 kb upstream from the transcription start site to 40 kb downstream

443 of the transcription termination site demarcated the coding sequence. Finally, 28,275
444 SNPs in ALL-DILI, 27,740 SNPs in CM-DILI, and 28,069 SNPs in HC-DILI were
445 selected (PRS_{gw}).

446 Additionally, we developed CM-DILI PRS_{gw+} from CM-DILI GWAS summary
447 statistics expanding out data set to 928 European CM-DILI cases due to inclusion of
448 157 flucloxacillin and 320 amoxicillin-clavulanate DILI patients which had not been
449 included in the GWAS sets for the CM-DILI PRS_{gw} [12]. For accurate evaluation of
450 TAK-875 DILI, we used 87,424 SNPs which were clumped by European LD after
451 filtering for imputation accuracy (Rsq ≥0.7) in TAK-875 GWAS sets to maximize
452 predictive accuracy for TAK-875 GWAS datasets. Next, in scoring PHH and iPSC, we
453 used 40,396 SNPs which were clumped by the LD after selecting Hapmap3 SNPs,
454 because the discovery GWAS used a reference panel (Haplotype Reference
455 Consortium) that was different from our available LD dataset (1KG European).

456

457 **GWAS enrichment analysis**

458 Enrichment analysis was performed using processed iDILIC/DILIN GWAS
459 summary statistics. GWAS analysis of regulatory or functional information enrichment
460 with LD correction (GARFIELD)[8] was performed using the default annotation data of the

461 software, using R package *garfield* v1.0.2 under R 3.3.3. Pathway scoring algorithm
462 (Pascal), integrating GWAS signals from multiple SNPs[9], was performed using Pascal
463 software (downloaded on 6, Jun 2017 from
464 https://www2.unil.ch/cbg/index.php?title=Pascal). For Pascal analysis, we used the
465 Molecular Signature Database (MSigDB) v6 gene sets[36,37] as pathway information and
466 default parameters of Pascal: maximum 3,000 SNPs were assigned to a gene and their
467 average GWAS signal was used for gene scoring, followed by pathway empirical
468 p-values obtained by their aggregation, considering LD in the European population.

469

470 **Cell cultures**

471 PHH were obtained from vendors (Lonza, Basel, Switzerland; KAC, Kyoto, Japan;
472 Sekisui Medical, Tokyo, Japan) listed in Supplementary Table S4 and cultured in
473 Hepatocyte Plating Medium (Lonza). Hepatocyte Culture Media (Lonza) was used for
474 thawing and daily sub-culturing. PHH were plated on Lumox plates (Sarstedt,
475 Nümbrecht, Germany) coated with 5 μg/cm^2 rat tail type-I collagen (Corning, Corning,
476 NY, USA) in 0.02 mol/L acetic acid. PHH were plated at 4×10^5 cells per well on a
477 24-well-type Lumox multiwell plate (Sarstedt) for DNA and RNA extraction and 5×10^4
478 cells per well on a 96-well-type Lumox multiwell plate for drug sensitivity assays. After

479 24 h of preculture, cells were treated with LCA or 1% DMSO for 1 h. After washing with
480 PBS, cells were treated with the indicated drugs for viability assay and transcriptome
481 analysis.

482 A human iPSC line, 1383D2, was kindly provided by the Center for iPS Cell
483 Research and Application, Kyoto University. Other human iPSC lines (CW10027, 73-B,
484 75-A, 10-A, 45-A) were obtained from Coriell institute (NJ, USA) or European Bank for
485 Induced Pluripotent Stem Cells (EBiSC) (Cambridge, UK). All iPSCs used herein had
486 normal karyotypes (data not shown). All iPSCs were cultured on laminin 511
487 E8-fragment-coated (iMatrix-511; Nippi, Tokyo, Japan) dishes in StemFit AK02N
488 (Ajinomoto, Tokyo, Japan). Directed differentiation methods for each lineage are
489 described previously[13]. To generate iPSC-HLOs *in vitro*, 1.8×10^6 cells per microwell at a
490 ratio of 10:7:1 (human iPSC-HE/iPSC-EC/iPSC-STM) were resuspended in a mixture of
491 endothelial cell growth medium (EGM, Lonza) and HCM containing dexamethasone
492 (0.1 mM; Sigma-Aldrich, St. Louis, MO, USA), oncostatin M (10 ng/mL; R&D Systems,
493 Minneapolis, MN, USA), hepatocyte growth factor (HGF) (20 ng/mL, PromoKine,
494 Heidelberg, Germany), and SingleQuots (Lonza) and plated on 6-well plate of the
495 Elplasia micro-space cell culture plate (Kuraray, Tokyo, Japan). After 24 h, the
496 generated iPSC-HLOs were harvested via gentle pipetting and transferred into a

single-use bioreactor culture system (30 mL; ABLE, Tokyo, Japan) and cultured in the same media. Further, after differentiation induction for over 10 days, viability testing and transcriptome after drug addition were performed. Viability testing and transcriptome were performed on iPSC-HLOs after hepatic differentiation for 10 days and more.

Drug treatments

DILI drug compounds were dissolved in DMSO to a final concentration of 1% in cell culture medium. Each drug was exposed in 3 doses for 24 or 72 hours with or without LCA 1 hour pre-treatment. The following DILI drugs used in this article categorized into 2 different DILI; cholestatic/mixed (CM)-DILI drugs: cyclosporin A (Wako), carbamazepine (Tokyo Chemical Industry), ketoconazole (Sigma), troglitazone (Wako), bosentan (Toronto Research Chemicals), flutamide (LKT laboratories), diclofenac (Wako), amoxicillin-clavulanate (amoxicillin; Wako, clavulanate; Matrix Scientific) and methapyrilene (Toronto Research Chemicals); hepatocellular(HC)-DILI drugs: tacrine (Cayman), tolcapone (Toris Bioscience) and acetaminophen (Toronto Research Chemicals).

Transcriptome analysis

Total RNA was extracted from cultured cells, using a PureLink RNA Mini Kit

(Thermo Fisher Scientific). A cDNA library was generated from 10 ng of total RNA, using a SuperScript VILO cDNA Synthesis Kit (Thermo Fisher Scientific). Amplification, primer digestion, and adapter ligation were performed using an Ion AmpliSeq Transcriptome Human Gene Expression Kit (Thermo Fisher Scientific). The cDNA library was purified using Agencourt AMPure XP Reagent (Beckman Coulter) and quantified using Ion Library TaqMan Quantitation Kit (Thermo Fisher Scientific), followed by dilution to 75 pM with water and pooled equally. Eight samples per pool were sequenced using an Ion 540 Chip Kit (Thermo Fisher Scientific) simultaneously, using IonS5 XL (Life technologies) and Ion Chef instrument systems (Life technologies) with Ion 540 Kit-Chef (Life technologies). All procedures were performed per the manufacturers' protocols. From the sequencing output, quality control, alignment reads on hg19 (hg19_AmpliSeq_Transcriptome_21K_v1.bed), read counts and the normalization (reads per million, RPM) for each gene were obtained using the Torrent ampliSeqRNA plugin (hg19 AmpliSeq Transcriptome ERCC v1) in Torrent Suite Software v5.2.1.

GSEA for PHH models was performed using GSEA software (v2.2.3, Broad Institute) in the prerank mode[36], using previously described gene sets. The permutation time of GSEA was set to 10,000 with other default parameters.

In-house whole-genome genotyping and PRS calculation

DNA from iPSC and PHH cell was extracted using PureLink Genomic DNA Mini Kit (Thermo Fisher Scientific) per the manufacturer's protocols. These subjects were genotyped using Infinium OmniExpressExome-8 v1.4 BeadChip (Illumina). Corresponding manifest files of the genotyping array were obtained from Illumina. QC analysis, haplotype phasing, and genotype imputation the genotyped data were performed along with the TAK-875 GWAS. Notably, Hardy–Weinberg equilibrium was not assessed owing to the small sample size; outliers of the known-ancestry cluster were assessed via PCA for SNPs, using Hapmap3 release3[38]. After genotype imputation using 1KGP3 reference panel, genotype dosage of variants with a minimac3 imputation quality (Rsq)≥0.7 in a batch was used for calculating PRS similar to previously described criteria.

PHH viability assay

PHH viability was determined using a CellTiter-Glo luminescent cell viability assay kit (Promega) per the manufacturer's protocol. After pre-treatment with LCA (Sigma-Aldrich), relative cell viability was determined as a ratio between cell viability of treated LCA-pre-treated samples and that of control LCA-pre-treated samples. Viability

upon LCA treatment was calculated as the ratio of cell viability of control LCA-pre-treated samples and that of control non-pre-treated (1% DMSO treated) samples. Cell viability was determined for each experimental batch (n=3 in a condition; n=1-3 batches from available cryopreserved PHH vials).

iPSC-HLOs viability assay

iPSC-HLO cell viability after drug treatment was determined using a CellTiter-Glo 3D Cell Viability Assay (Promega) per the manufacturer's protocol. For drug evaluation, over 100 iPSC-HLOs were dispensed per one well of low attached 96 well plate. Raw luminescence data were normalized via bright-field images with Cell^{3}iMager duos (SCREEN Holdings Co., Ltd., Kyoto, Japan) and dividing luminescence values in the area.

Live imaging

Indicated molecules or proteins were stained using live cell imaging kits or antibodies, per the manufacturer's instructions, as follows: dead cells, LIVE/DEAD Cell Imaging Kit(488/570) (Thermo); bile acid transport, CLF (BD Biosciences Discovery

Labware, Franklin Lakes, NJ, USA); reactive oxygen species, CellROX green reagents (Thermo Fisher Scientific).

Immunostaining

iPSC-HLOs were fixed in 4% paraformaldehyde (Wako) in PBS for 15 min. Samples were blocked with Donkey serum (Millipore) and probed with primary antibodies against Albumin (SIGMA), BSEP (SIGMA), CD31 (BD), CHOP (Santa-Cruz), and HNF4a (Santa-Cruz) at 4°C overnight. Samples were probed with secondary antibodies conjugated with Alexa Fluor (Life Technologies) and DAPI (Sigma) for nuclear staining. Images were acquired using LSM 880 with Airy scan (Zeiss).

Immunoblot

Drug-treated iPSC-HLOs were suspended in Laemmli sample buffer (BioRad). Proteins were resolved via SDS-PAGE and electro-transferred onto PVDF membranes (BioRad). Western blotting was performed per standard protocols, and blotted protein samples were blocked with Blocking One (Nakarai) for 1 h. Samples were incubated with anti-KDEL (Enzo), anti-XBP1-s (Cell Signaling), and anti-GAPDH (Cell Signaling)

primary antibodies at 4°C overnight and probed with horseradish peroxidase-conjugated secondary IgG (Cell Signaling) for 1 h. Signals enhanced by ECL Prime (GE Healthcare) were detected using ChemiDoc Touch imaging system (BioRad). Protein expression levels were measured by Image Lab software (BioRad).

Enzyme-linked immunosorbent assay (ELISA)

Human albumin (ALB) in was quantified using a Human Albumin ELISA Quantitation Kit (Bethyl Laboratories Inc., Montgomery, TX, USA) per the manufacturer's protocols.

CYP3A4 activity

Cell-based CYP3A4 activity of iPSC-HLOs was measured using P450-Glo™ CYP3A4 Assay (Promega) per the manufacturer's protocols. Raw luminescence data were normalised using bright-field images with Cell^{3}iMager duos and dividing luminescence values in the area.

Raw luminescence data were normalized to cell numbers by dividing the P450-Glo Assay values by CellTiter-Glo 3D Cell Viability Assay (Promega).

.

GSH/ GSSG ratio

Changes in GSH/GSSG, oxidative stress indicator, in iPSC-HLOs under drug treatment was quantified via a luminescence-based GSH/GSSG-Glo assay (Promega) per the manufacturer's instructions.

Transcriptome analysis of multi donor liver tissues

Genotypes and transcriptome data were obtained from ref#[16] via https://www.synapse.org/#!Synapse:syn4492. From curated SNP array dataset including 349,085 variants and 195 donors, we further performed the following quality controls: exclude variants with genotype call rate < 95%, Hardy P -value < 10^{-6}, and A/T or G/C SNPs; exclude individuals in non-European ancestry. From the remaining 213,257 variants and 161 donors, we phased and imputed variants as described above and calculated their PRS. From DNA microarray-based transcriptome datasets of liver tissue (40,638 transcripts and 467 donors), we excluded >10% missing donors and ≥1 missing transcripts and aggregated the probe-level datasets into gene-level datasets by replacing with their average (25,015 transcripts and 157 donors remained). From the overlapped 156 donors, we calculated coefficient effect of PRS_{gw} on each transcript by

620 regressing out the five first transcriptomic PCs to exclude the effects of batch effects in
621 the transcriptome datasets. Preranked GSEA analysis on Reactome genesets was
622 performed by using the ranking of their t-statistics.

623

624 **Re-analysis of TG-GATEs**

625 Transcriptome data of PHH under drug treatments (2,605 data points including 158
626 types of drugs, three types of concentration, and three types of exposure times)
627 measured by GeneChip Human Genome U133 Plus 2.0 array (Affymetrix) was obtained
628 from the TG-GATEs website (https://toxico.nibiohn.go.jp/english/index.html)[18]. We
629 carried out normalization of expression levels using Frozen Robust Multiarray Analysis
630 using the R package 'frma' (version 1.14.0)[39] with default parameters and took average
631 expression levels within replicates. We calculated a signal intensity ratio to each control
632 data for each probe, took an average value among probe sets for each gene, and made
633 a gene rank matrix for preranked GSEA analysis.

634

635 **Data availability**

636 The AmpliSeq data used herein are deposited in GEO under accession number
637 GSE152447. The genotype data for TAK-875 DILI GWAS, the related phenotype

information, and the summary statistics are stored in Takeda Pharmaceutical Company Ltd, due to the ethical approval in this study. These datasets are available upon reasonable request to Takeda Pharmaceutical Company Ltd via the corresponding author and after being approved by the Ethics Committee of Takeda Pharmaceutical Company Ltd. Web-links of publicly available datasets are as follows: transcriptome data of PHH under drug treatments (TG-GATEs), https://toxico.nibiohn.go.jp/english/index.html; transcriptome data of multi-donor liver tissues, https://www.synapse.org/#!Synapse:syn4492; transcriptome data of PHH (SRR4000958), https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR4000958; GENCODE (v26); https://www.gencodegenes.org; MSigDB (v6), https://www.gsea-msigdb.org/gsea/msigdb/index.jsp; 1KGP3 imputation reference panel: https://genome.sph.umich.edu/wiki/Minimac3.

Code availability

We used publicly available software and parameters as described in this Methods section for the analysis. The software programs are available from the following URLs: plink, https://www.cog-genomics.org/plink2/; Minimac3, https://genome.sph.umich.edu/wiki/Minimac3; snptest,

https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html; FaQCs, https://github.com/chienchi/FaQCs; PRSice, http://PRSice.info; STAR, https://github.com/alexdobin/STAR; cufflinks, http://cole-trapnell-lab.github.io/cufflinks/; GARFIELD, https://www.ebi.ac.uk/birney-srv/GARFIELD/; Pascal; https://www2.unil.ch/cbg/index.php?title=Pascal; GSEA, https://www.gsea-msigdb.org/gsea/; R, https://cran.r-project.org/.

664 **Methods only References**

665

666 26. Gibbs, R. A. *et al.* A global reference for human genetic variation. *Nature* **526**,
667 68–74 (2015).

668 27. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger
669 and richer datasets. *Gigascience* **4**, 7 (2015).

670 28. Loh, P.-R. *et al.* Reference-based phasing using the Haplotype Reference
671 Consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016).

672 29. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat.*
673 *Genet.* **48**, 1284–1287 (2016).

674 30. Marchini, J. & Howie, B. Genotype imputation for genome-wide association
675 studies. *Nat. Rev. Genet.* **11**, 499–511 (2010).

676 31. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: Polygenic Risk Score
677 software. *Bioinformatics* **31**, 1466–1468 (2015).

678 32. Asai, A. *et al.* Paracrine signals regulate human liver organoid maturation from
679 induced pluripotent stem cells. *Development* **144**, 1056–1064 (2017).

680 33. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–

21 (2013).

34. Lo, C.-C. & Chain, P. S. G. Rapid evaluation and quality control of next generation sequencing data with FaQCs. *BMC Bioinformatics* **15**, 366 (2014).

35. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).

36. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).

37. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).

38. Frazer, K. A. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).

39. McCall, M. N., Bolstad, B. M. & Irizarry, R. A. Frozen robust multiarray analysis (fRMA). *Biostatistics* **11**, 242–253 (2010).

40. Fabregat, A. *et al.* The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* **46**, D649–D655 (2018).

Acknowledgments

We thank Shinya Yamanaka, Seigo Izumo and Yasushi Kajii for their critical comments. We thank Tamaki Kono, Kohei Araki, Kazuaki Enya and Hirokazu Kawaguchi for technical and analytical supports and Wendy L. Thompson for critical reading of the manuscript. Authors thank the Drug Induced Liver Injury Network (DILIN) and the international Drug-Induced Liver Injury Consortium (iDILIC) for providing its data included in this paper. The DILIN and iDILIC were not involved in data analyses, manuscript preparation or its review. This study was supported by T-CiRA Joint Program from Takeda Pharmaceutical Co., Ltd. to T.T.. T.T. is a New York Stem Cell Foundation – Robertson Investigator and also recipient of Cincinnati Children's Research Foundation grant, NIH grant UG3 DK119982, Dr. Ralph and Marian Falk Medical Research Trust Awards Program, Takeda Science Foundation award, Mitsubishi Foundation award, AMED JP19fk0210037, JP19bm0704025, JP19fk0210060, JP19bm0404045, and JSPS JP18H02800, 19K22416.

Author Contributions

M.K., E.K., and T.T. conceived and designed the study, analyzed the data and wrote the

manuscript; P.N., A.K.D., G.P.A. and P.B.W. conducted iDILIC/DILIN DILI GWAS and critically revised the manuscript. M.K., M.O., E.K. and T.T. conducted the other data analysis; E.K. J.F., Y.Noguchi. performed cell culture experiments. Y.Nio., T.S, H.A., Y.D., and T.T. provided critical discussion.

Competing interests

T.T. received research funding related to this project from Takeda Pharmaceutical Co., Ltd.. M.O., E.K., Y.Nio., T.S, Y.D. and H.A. are employees of Takeda Pharmaceutical Co. Ltd. The remaining authors declare no competing interests. PN is an employee of Sema4.

Figures

Figure 1. Polygenic architecture of drug induced liver injury (DILI).

(a) Summary of PRS and their sources. **(b)** Manhattan plot of p-values (logistic regression model, from snptest software) in TAK-875 white GWAS. Blue line is a suggestive cut-off ($p<1\times10^{-5}$). **(c)** Polygenic test using CM-DILI GWAS summary statistics. X-axis, total number of SNPs, ordered by iDILIC/DILIN GWAS association; y-axis, explained variance of TAK-875 DILI phenotype from iDILIC/DILIN GWAS signals; colour scale, one-tail p-value for the shared genetic aetiology test from PRSice software. The described P_{GWAS} indicated the corresponding thresholds to x-axis. **(d)** GARFILD plot of CM-DILI GWAS summary statistics in Nicoletti et al., 2017. Enrichment in FAIRE-seq (Formaldehyde-Assisted Isolation of Regulatory Elements) annotations in *left* and genic state in *right*. **(e)** Enriched experimentally known or related pathways associated with CM-DILI. CM-DILI GWAS summary statistics in Nicoletti et al., 2017 were used. The empirical p-values were from Pascal software. **(f)** Distribution of PRS in TAK-875 DILI patients (cases, n=39 individuals) and matched patients treated with TAK-875 without DILI (controls, n=122 individuals). For each type of PRS, scores were centred on the median values in controls. *, $p < 0.05$ (two-tailed Wilcoxon–Mann–Whitney U test; p = 0.0036 for CM-PRS_{gw}, p = 0.70 for HC-PRS_{gw}, p = 0.27 for

ALL-PRS_{gw}, and p = 0.029 for CM-PRS_{gw+}). For the box plot, the box represented the first and third quartiles, the center line represented the median, the upper whisker extended from the hinge to the highest value that is within 1.5 × IQR (inter-quartile range) of the hinge, the lower whisker extended from the hinge to the lowest value within 1.5 × IQR of the hinge, and the data beyond the end of the whiskers were plotted as points.

Figure 2. Polygenic score based human stratification by in vitro multi-drug induced CM-DILI assays.

(**a**) Workflow of comparison. (**b**) Viability comparison of multi-donor iPSC-HLO models (triangle) and primary human hepatocytes (circle) under multiple DILI drug treatment. See Table S4 for donor information per drug

(**c**) Comparison of pearson correlation coefficient between the survival rate of the liver model (mean value for each donor) and the PRS (PRS_{gw} or PRS_{gw+}) under indicated drugs. *, p < 0.01 (two-tailed Welch's t-test); p = 0.0038 for cholestatic DILI drugs treatment and p = 0.0005 for hepatocellular DILI drugs treatment. For the violin plot, the center point represented the median, the upper whisker extended from the hinge to the highest value that is within 1.5 IQR (inter-quartile range) of the hinge, the lower

764 whisker extended from the hinge to the lowest value within 1.5 IQR of the hinge.

765

766 **Figure 3. Mechanistic association studies for CM-DILI vulnerability.**

767 **(a)** Method to find CM-DILI PRS_{gw}-associated pathways. We performed GSEA analysis
768 of CM-DILI PRS_{gw}-associated genes, regressing out five first transcriptome principal
769 components to capture experimental variability[15] **(b)** Pathway enrichment analysis
770 results for PHH cells and liver tissues. (**c**) Heatmap analysis of gene sets involved in
771 unfolded protein response (76 genes) and TCA cycle and respiratory electron transport
772 (116 genes) (Reactome pathway database[40]) in multi-donor iPSC-HLO models. (**d**)
773 Immunostaining of HNF4a and CHOP under the indicated CsA treatment. The
774 representative pictures from 2 biological replicates. (**e**) XBP1s and KDEL levels under
775 the indicated CsA treatment via immunoblot analysis. GAPDH was used for loading
776 controls. n=3, independent experiments. (**f**) Live imaging of oxidative stress induction
777 using CellROX reagent (ROX), CLF accumulation, and cell death (stained by PI) under
778 the indicated CsA treatment. The representative pictures from 3 biological replicates. (**g**)
779 Comparison of GSH/GSSG ratio in iPSC-HLO model between the PRS_{gw}-high donor
780 (45A) and the -low donor (CW10027) upon CsA treatment. * ; p=0.017064, a significant
781 difference between the two groups. n=3, independent experiments. (**h**) GSH/GSSG

ratio change by BM treatment in 1383D2 derived iPSC-HLOs under basal condition. n=2, independent experiments. (**i**) Cell viability upon BM treatment between PRS_{gw}-high and -low donor-derived iPSC-HLOs under CsA 50 μM treatment. Significant difference between the two groups was showed p = 0.0051 with no-treatment, whereas not shown p = 0.1048 with BM pre-treatment and p=0.0640 with BM pre-/co-treatment. n=3, independent experiments. All data was shown as mean±SD, *; $p < 0.05$ (two-tailed Welch's t-test).

Figure 4. Polygenic risk score associated transcriptomic signatures associate with known CM-DILI responses.

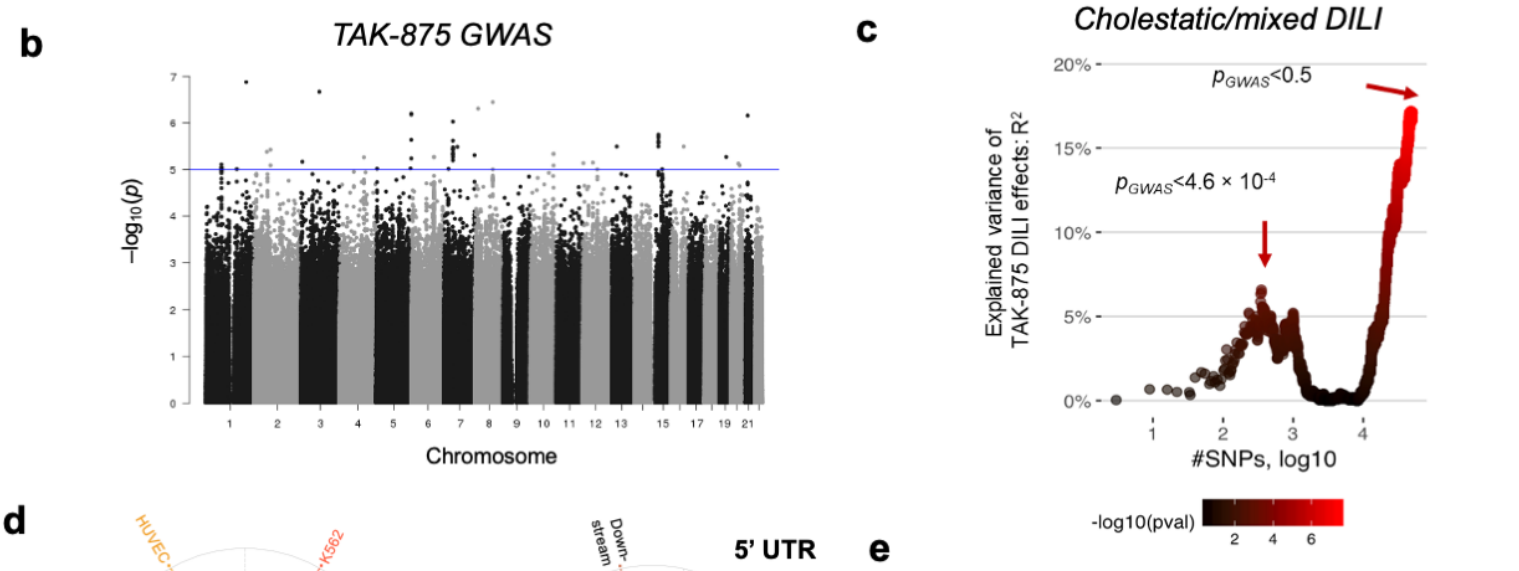
(a) A schematic representation of the protocol for identifying compounds associated with PRS_{gw} related transcriptomic pathways (see Results and Methods). **(b)** Heatmap analysis of gene sets involved in TCA cycle and respiratory electron transport in 8 PHH donor ordered by PRS_{gw}. This pathway was one of the significantly inactivated pathways in higher CM-PRS_{gw} donors (Supplementary Table S9). The core enriched genes in GSEA were shown. **(c)** Network representation of screened gene sets associated with CM-DILI PRS_{gw} (FDR < 0.01). Representative clusters and compound associated with them were indicated. All of the raw results were shown in

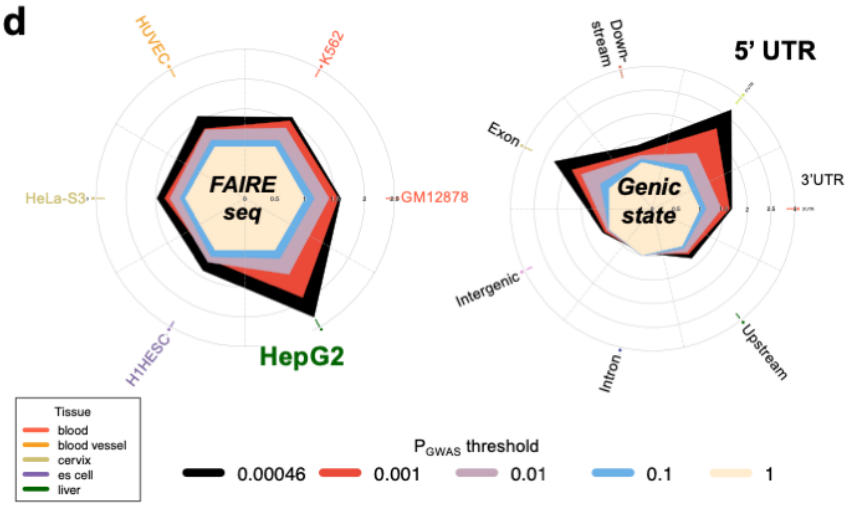
Supplementary Table S9 and S10. **(d)** Gene expression analysis for representative ER stress marker genes in the 8 PHH donor under CsA treatment or control condition (see Methods). **, $p < 0.01$ (two-sided Pearson correlation test; $p = 4.7\times10^{-3}$) **(e)** PRS_{gw} informed CM-DILI vulnerable mechanisms, genetic factors and their relationships. Green text, validated events by phenotypic assays; Red box, polygenic risk score informed mechanisms for CM-DILI.

Figure 1. Polygenic architecture of drug induced liver injury (DILI).

a

Score	Description	#case (Flucloxacillin / amoxicillin-clavulanate)	#ctrl	#SNPs in PRS	Ref
CM-PRS$_{gw}$	cholestatic/mixed type of DILI	323 (0)	10,588	27,740	Nicoletti et al., 2017
CM-PRS$_{gw+}$	cholestatic/mixed type of DILI	928 (157/320)	10,397	40,396	Cirulli et al., 2019
HC-PRS$_{gw}$	hepatocellular type of DILI	474 (0)	10,588	28,069	Nicoletti et al., 2017
All-PRS$_{gw}$	all type of DILI	862 (0)	10,588	28,275	Nicoletti et al., 2017





e

Gene Set enriched in GWAS	P-value
PKA-mediated phosphorylation of CREB (oxidant-induced survival)	6.43×10^{-3}
Activation of chaperone genes by XBP1s (unfolded protein response)	9.86×10^{-3}
Citric acid cycle TCA cycle (mitochondrial activity)	1.04×10^{-2}

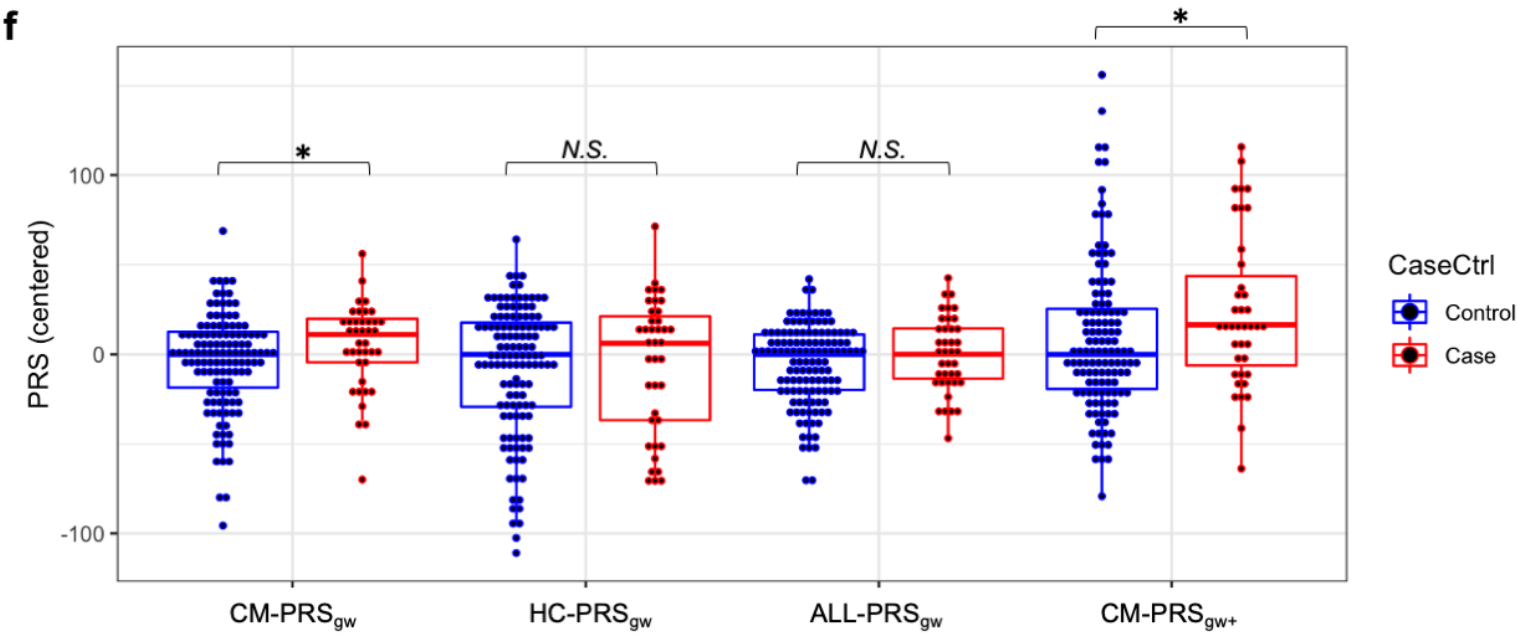


Figure 2. CM-DILI drugs-dependent DILI vulnerability difference from multi-donor assays.

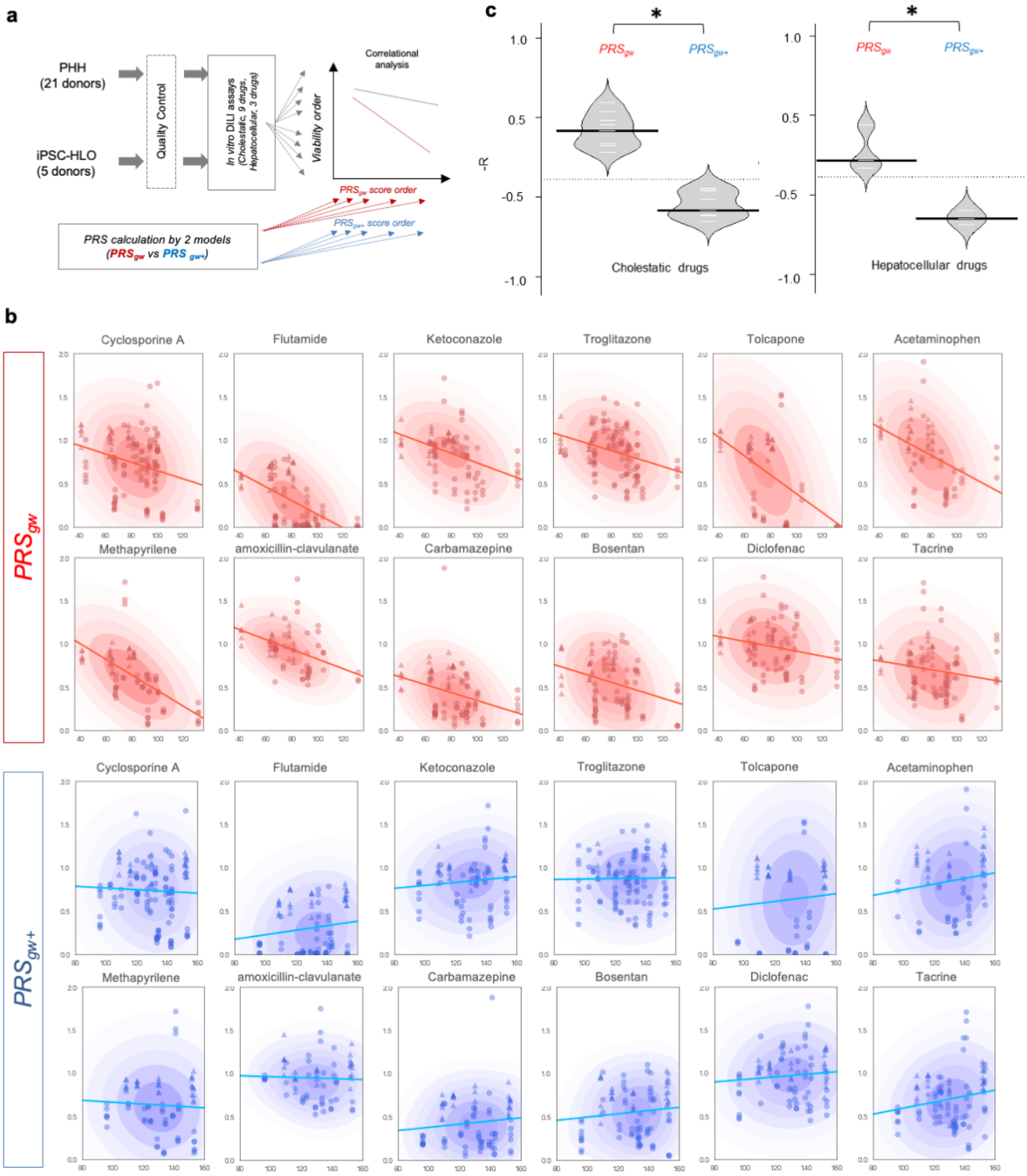


Figure 3. Mechanistic association studies for CM-DILI vulnerability.

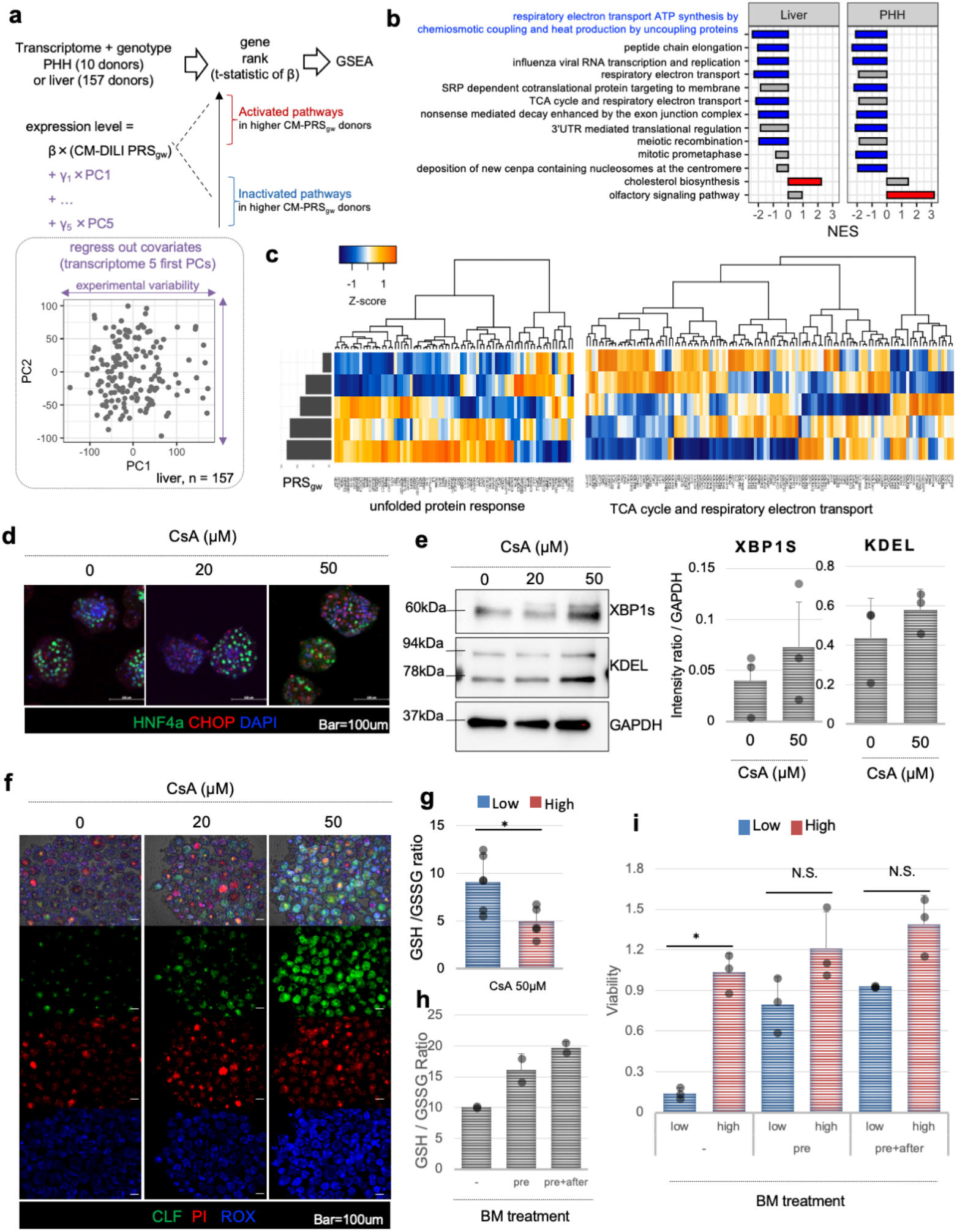
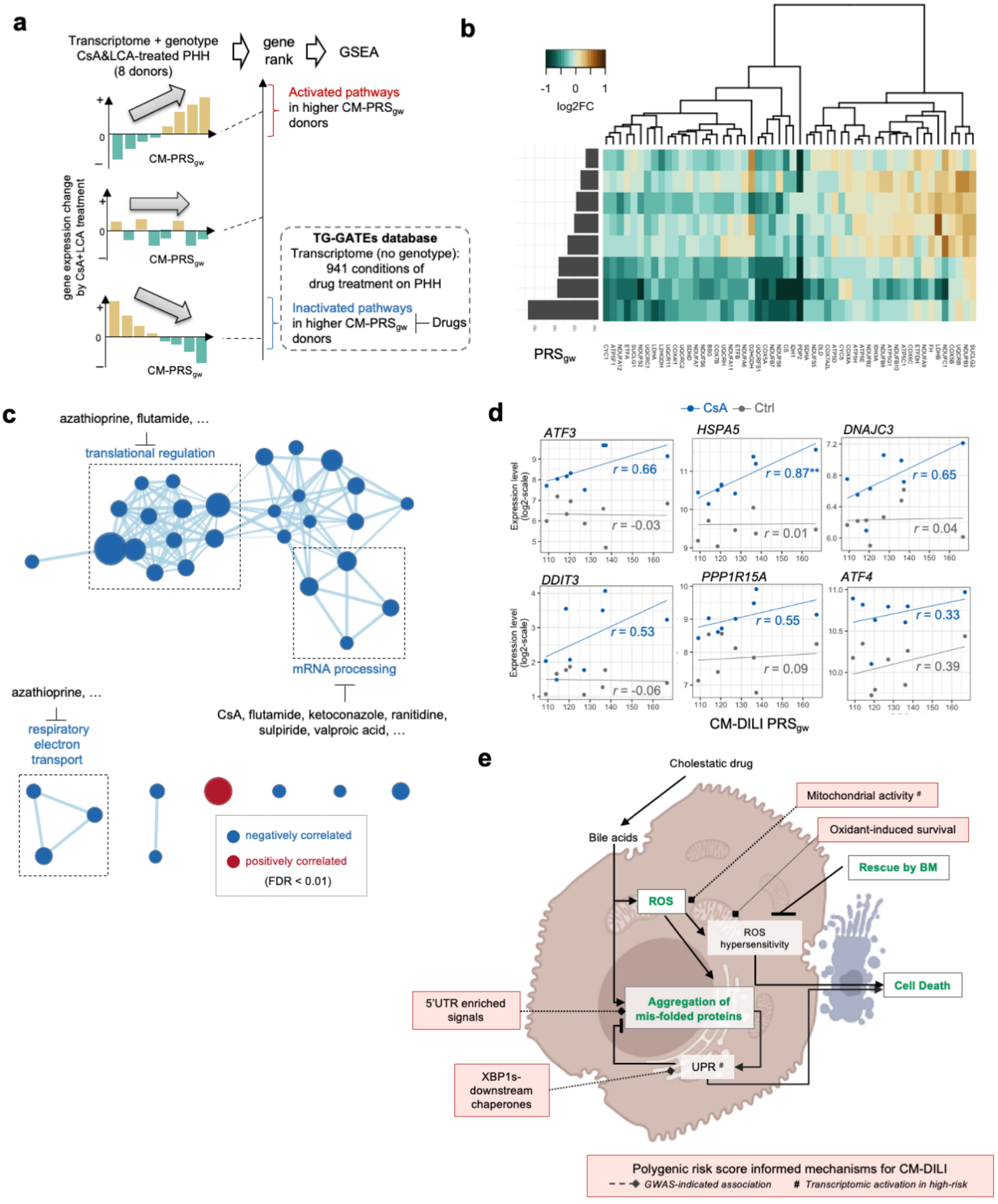
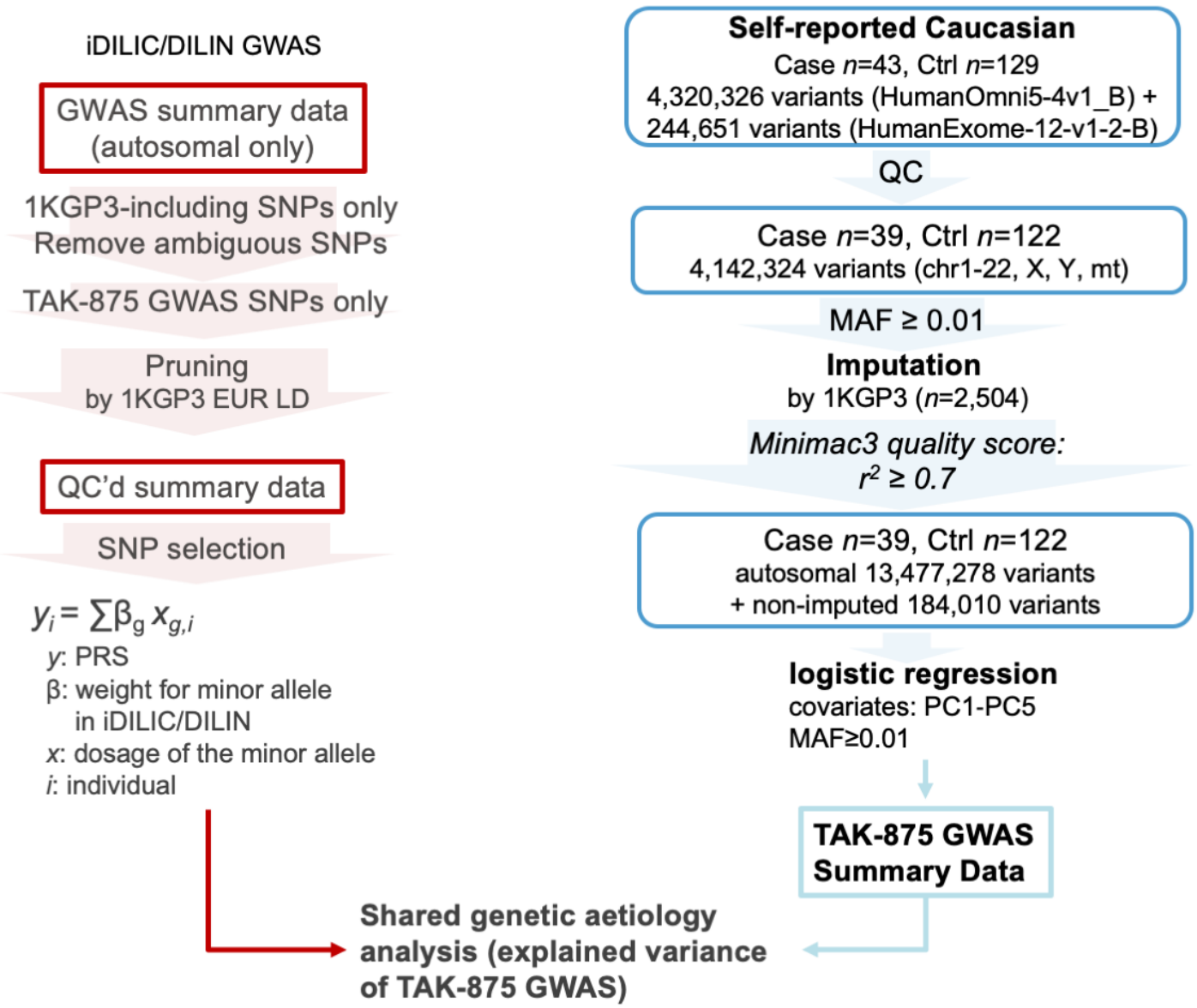


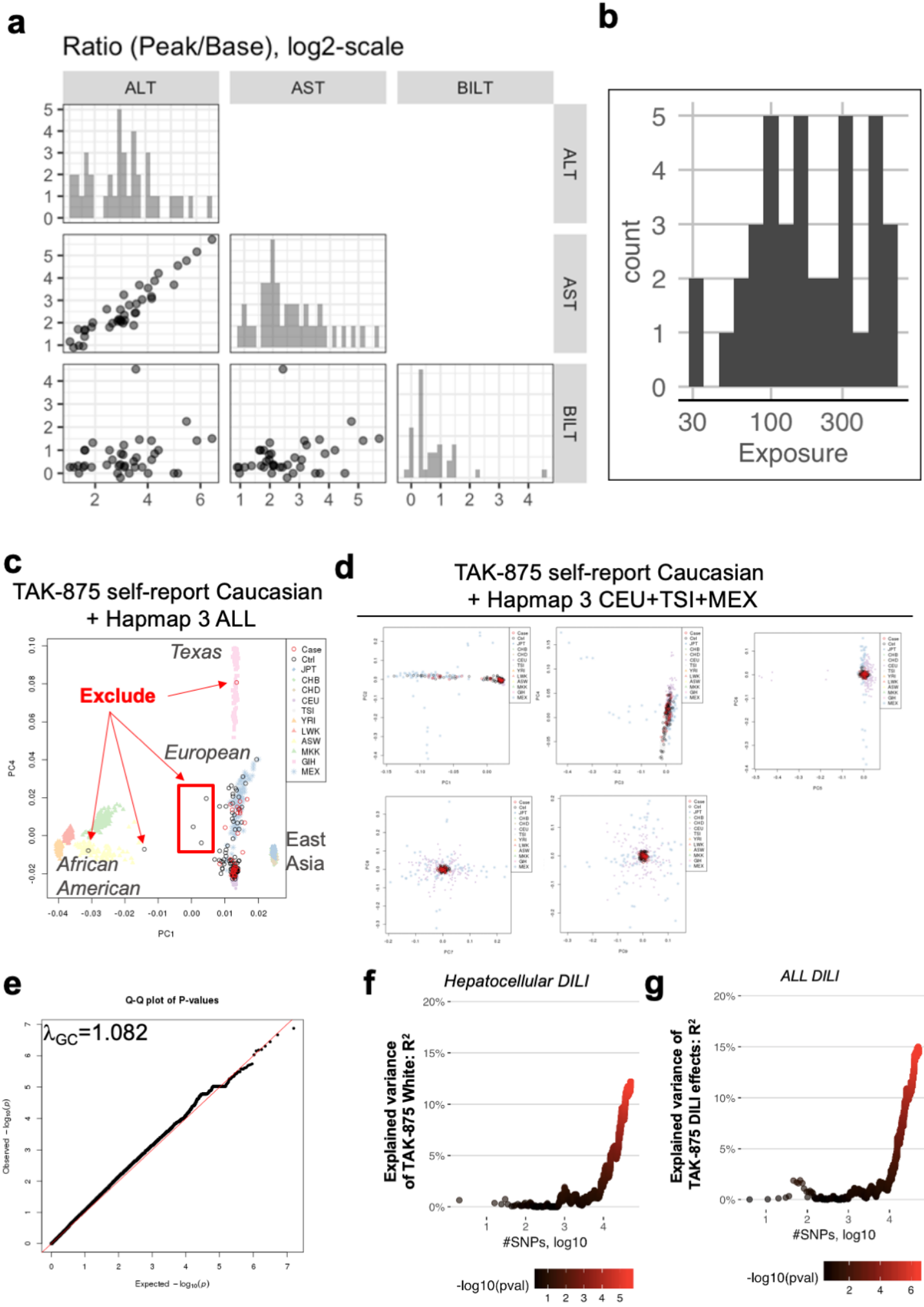
Figure 4. Polygenic risk score associated transcriptomic signatures associate with known CM-DILI responses.



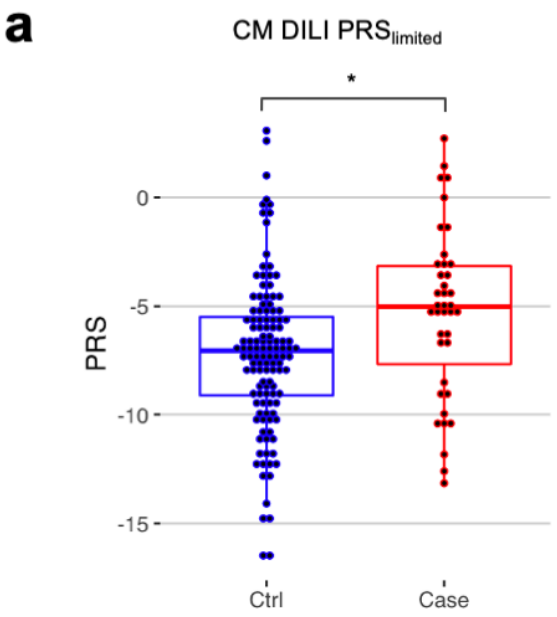
Extended Data Fig.1. Overview of our polygenicity analysis.



Extended Data Fig.2. TAK-875 DILI severity and GWAS analysis

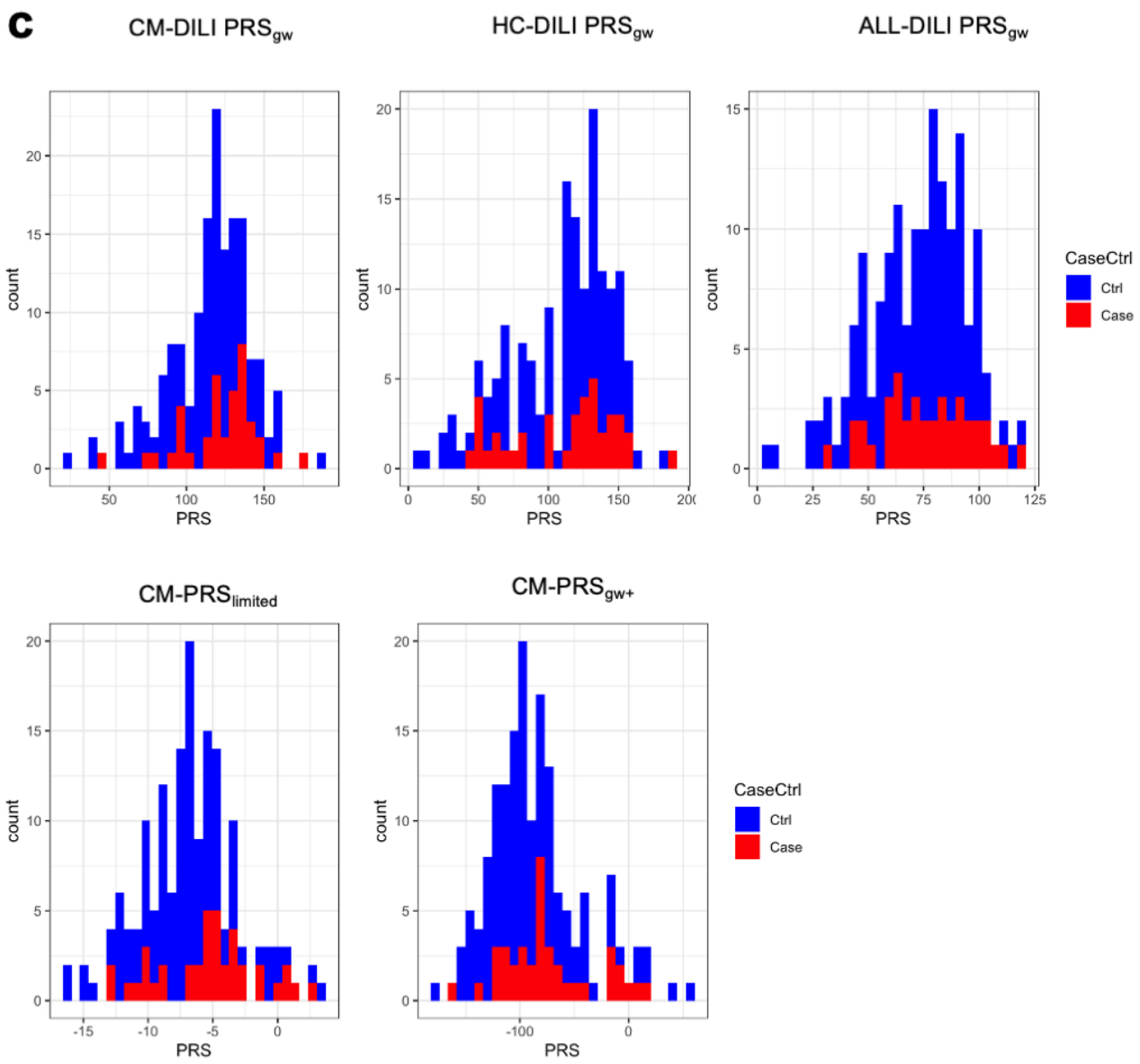


Extended Data Fig.3. Distribution and performance of each PRS

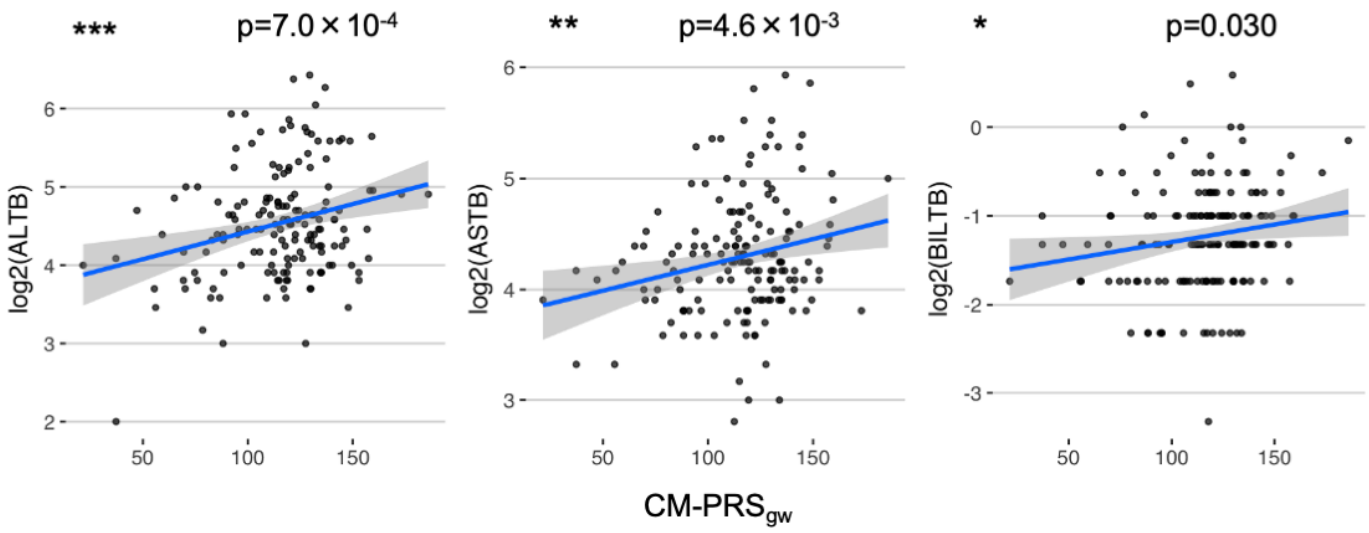


b

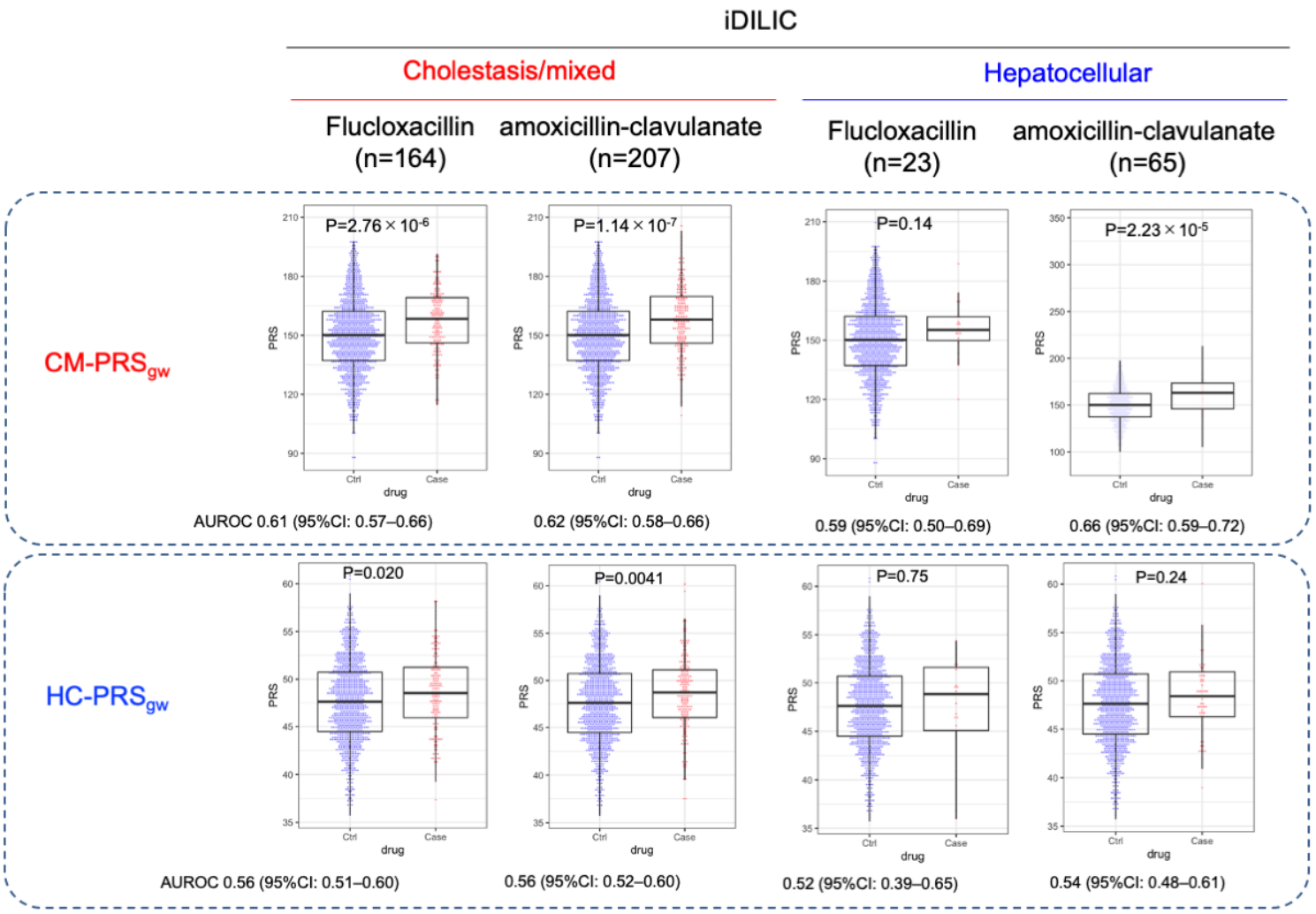
PRS type	AUROC (95% CI)
CM-PRS$_{gw+}$	0.62 (0.52–0.72)
CM-PRS$_{gw}$	0.61 (0.51–0.71)
CM–PRS$_{limited}$	0.67 (0.56–0.78)
HC-PRS$_{gw}$	0.52 (0.41–0.63)
All-PRS$_{gw}$	0.56 (0.45–0.66)



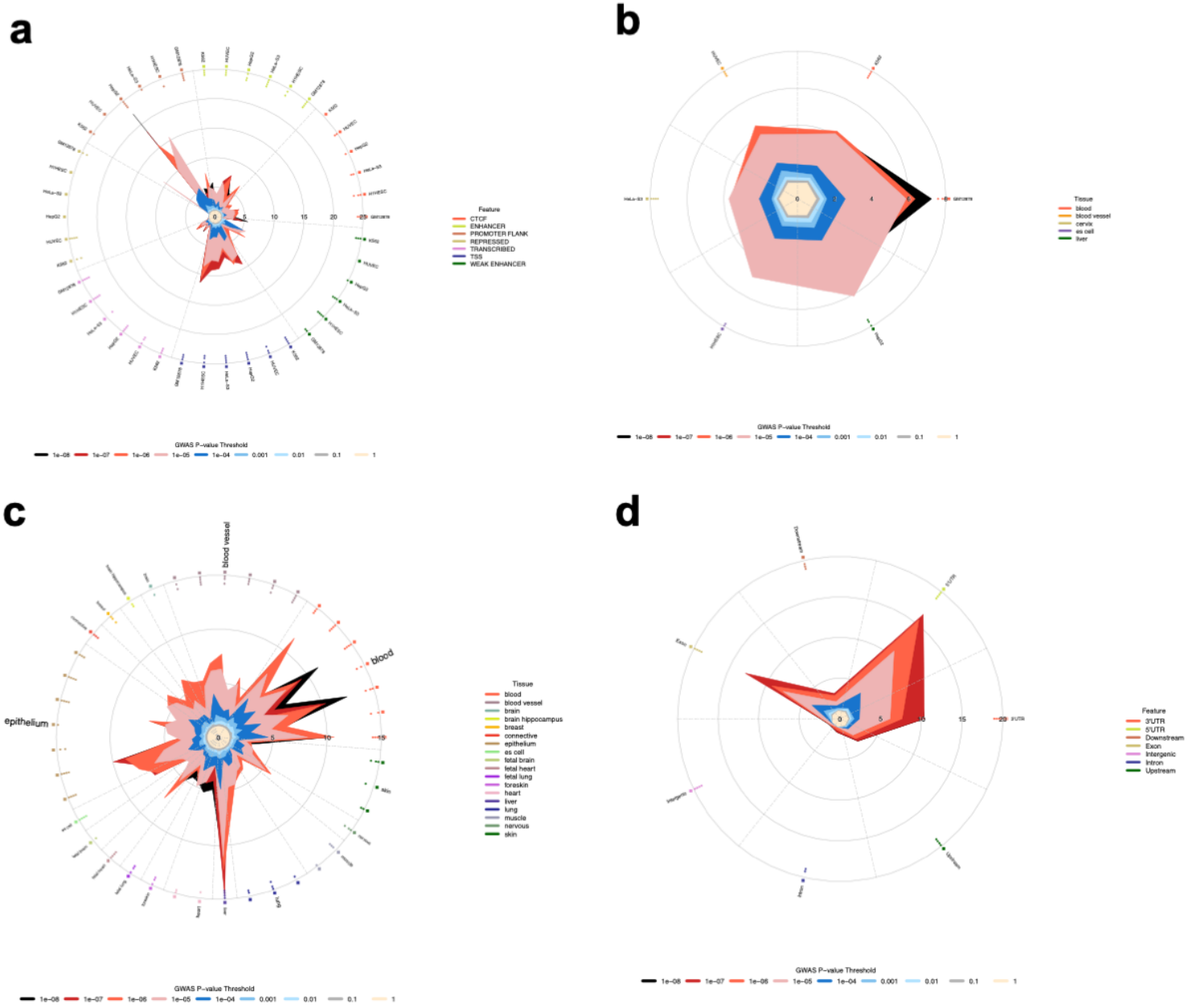
Extended Data Fig.4. Correlation between CM-PRS$_{gw}$ and biomarkers for DILI in TAK-875 treated subjects.



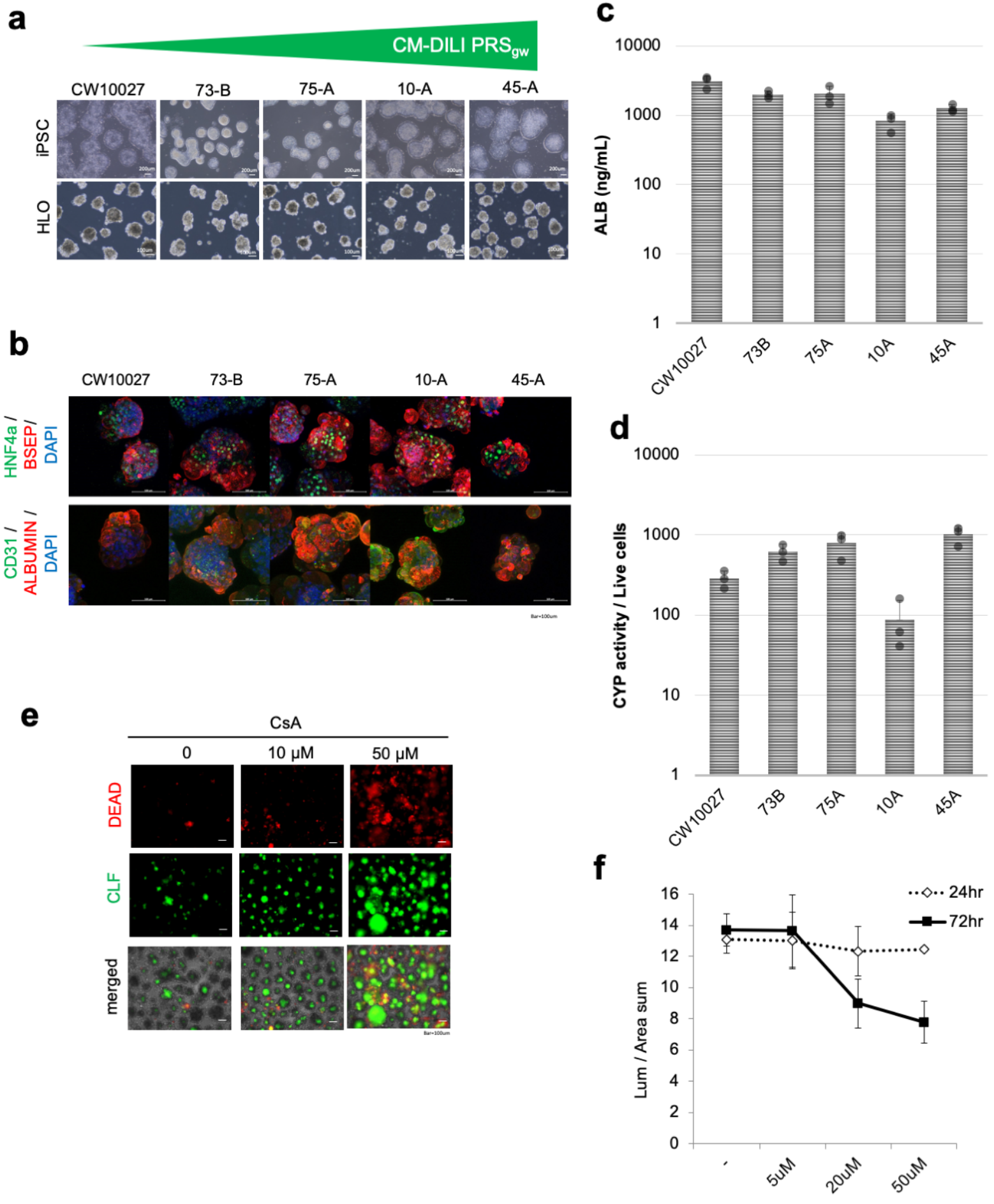
Extended Data Fig.5. Predictive accuracy of CM-DILI PRSgw for Flucloxacillin or amoxicillin-clavulanate DILI



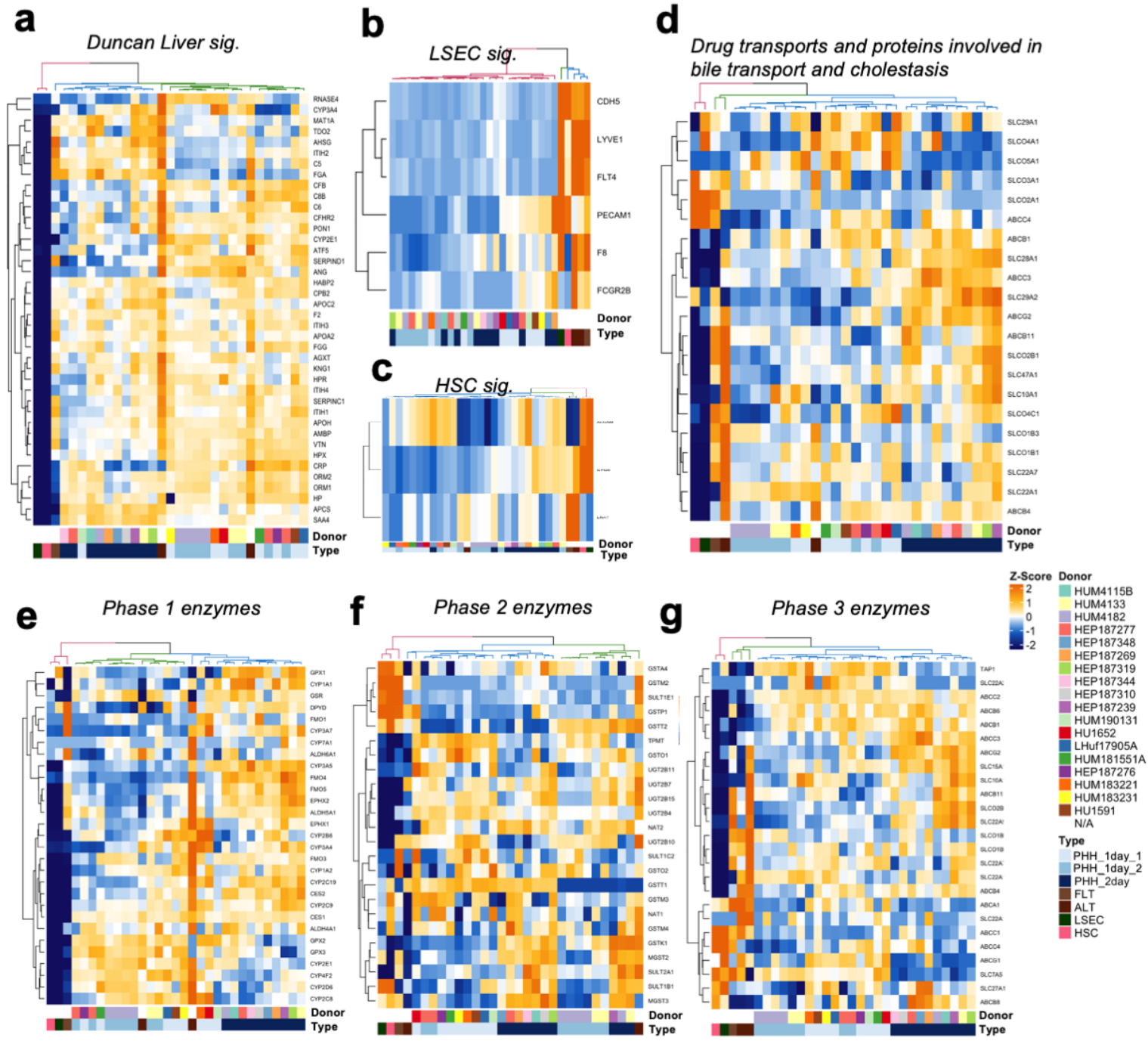
Extended Data Fig.6. GARFILD plot of CM-DILI GWAS summary statistics in Cirulli et al., 2019.



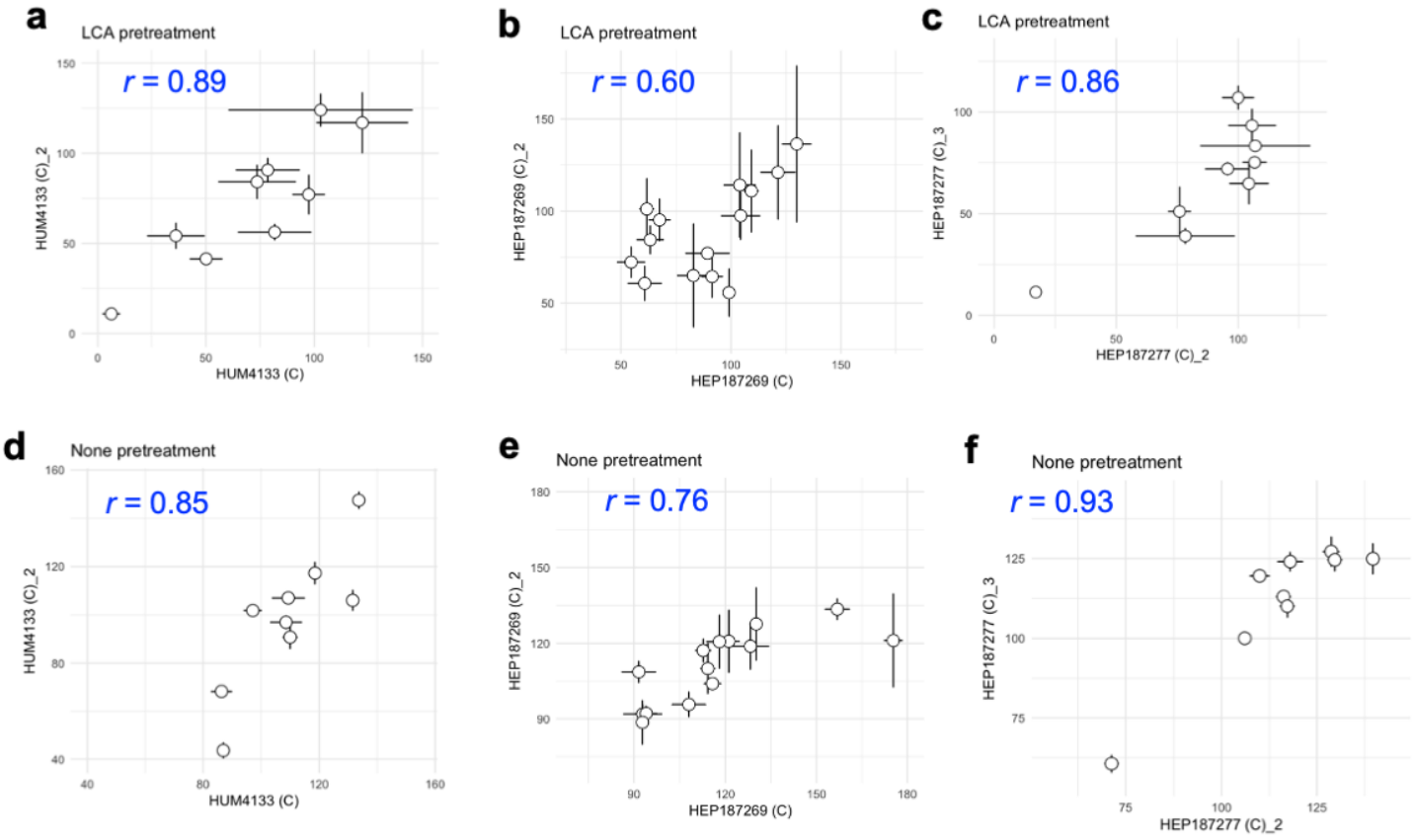
Extended Data S7. Multi-donor iPSC-HLO cholestatic DILI assays

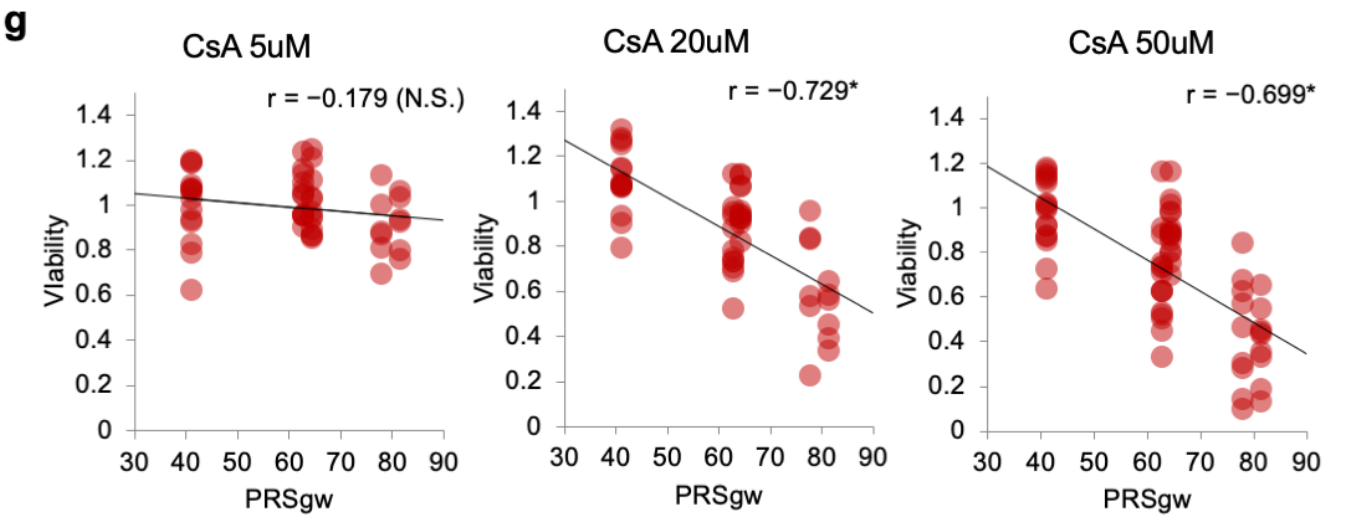


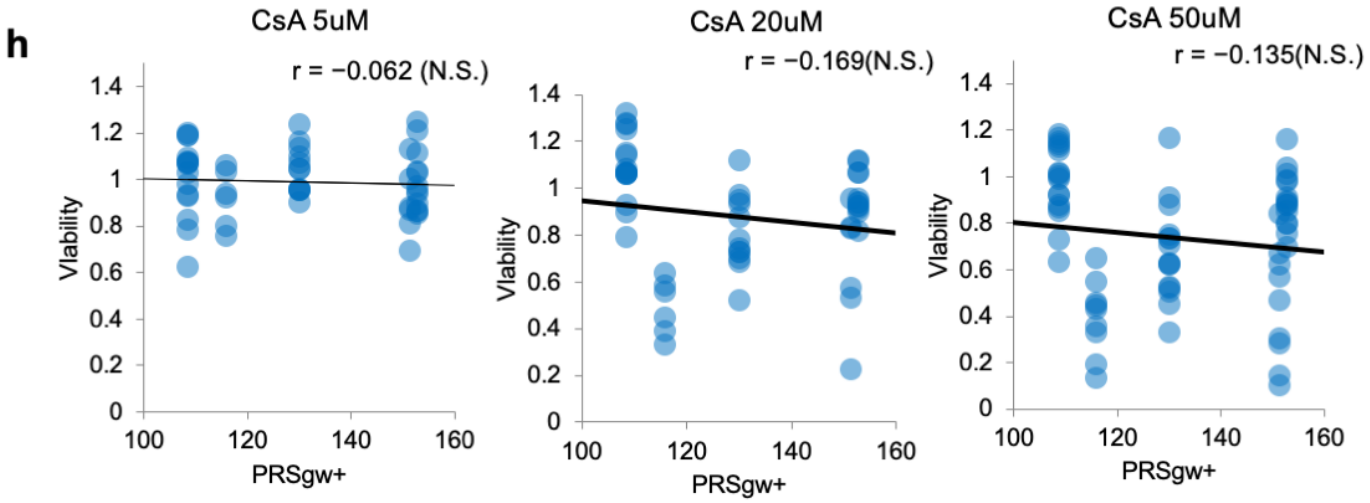
Extended Data Fig.8. Transcriptomic expression profiling of PHH at basal state



Extended Data Fig.9. Reproducibility of drug toxicity assay of PHH under LCA pretreatment







Extended Data Fig.10. Decreased mitochondria activity by cholestatic DILI drug treatment